DISSERTATION

EPSPS GENE DUPLICATION IN PALMER AMARANTH: RELATIVE FITNESS, INHERITANCE, AND DUPLICATION MECHANISM OF THE GLYPHOSATE RESISTANCE TRAIT

Submitted by

Darci A. Giacomini

Department of Bioagricultural Science and Pest Management

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2015

Doctoral Committee:

Advisor: Philip Westra
Co-Advisor: Sarah Ward

Jan Leach
Christopher Preston
Patricia Bedinger

Copyright by Darci A. Giacomini 2015

All Rights Reserved

ABSTRACT

EPSPS GENE DUPLICATION IN PALMER AMARANTH: RELATIVE FITNESS, INHERITANCE, AND DUPLICATION MECHANISM OF THE GLYPHOSATE RESISTANCE TRAIT

Glyphosate resistant (GR) Palmer amaranth (*Amaranthus palmeri* S. Wats.) is a weedy plant species that has invaded agricultural fields in at least 25 states, raising the cost of weed control to more than 4x the original cost. In most areas, the resistance is conferred through a gene amplification mechanism in which the target gene of glyphosate, 5-enolpyruvylshikimate-3-phosphate (*EPSPS*) is duplicated in the genome 100+ times, resulting in an overproduction of the EPSPS protein. With so much EPSPS enzyme available in each cell, glyphosate only inhibits a fraction of the proteins, leaving the rest to function as normal and ensuring plant survival. Understanding how this increase in *EPSPS* gene copy number and EPSPS protein production impacts relative fitness of the resistant plants was one objective of this research. Through greenhouse studies comparing high *EPSPS* copy GR plants with single copy sensitive plants, no difference was observed for any of the fitness characteristics measured. Both biotypes yielded similar numbers of offspring with no significant differences in germination or growth rate, revealing a complete lack of a fitness cost associated with the resistance trait. The second objective of this research was to quantify the stability of this resistance trait via multigenerational inheritance studies and within-plant *EPSPS* copy number variance measurements in the absence of glyphosate selection. The inheritance work found a complex pattern of *EPSPS* copy number transmission through the generations, a result that could be explained at least partially by

the mosaic of *EPSPS* gene copy numbers patterns observed in both male and female Palmer amaranth plants. Copy numbers were inherited in a non-Mendelian pattern with transgressive segregation of the trait seen in both directions (more and fewer *EPSPS* copies found in the offspring than expected). This retention of high *EPSPS* copy number in the absence of a glyphosate selection pressure and no evidence of a fitness cost associated with the resistance trait possibly indicates a long-term loss of glyphosate as a control option in fields infested with GR Palmer amaranth. The last objective of this project was to better understand the mechanism of *EPSPS* gene duplication through sequence assembly of the *EPSPS* amplicon and chromosomal localization of this duplicated region. The amplicon was extended out to a little over 110kb and was found to contain mostly repetitive sequence including long direct repeats, microsatellites, and multiple transposable elements. A fluorescent in situ hybridization (FISH) assay found a single chromosomal location for the *EPSPS* genes, suggesting a tandem gene arrangement. These results further suggest that *EPSPS* duplication is achieved in Palmer amaranth via unequal recombination of the repeats surrounding the gene during mitosis and/or meiosis.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Doug Sammons for being my champion through this entire process. Dr. Sammons has never failed to be a source of brilliant ideas, constant enthusiasm, and wholehearted support. His efforts to find me a spot within his group at Monsanto gave me the chance to work with him and his team members, all of whom were immensely knowledgeable and genuinely kind.

I'd like to express my deepest gratitude to my advisor, Dr. Phil Westra, for giving me an environment to excel as a scientist and grow as a person. I would not be where I am today without his unfailing support, guidance, and good humor. The same gratitude is extended to my co-advisor, Dr. Sarah Ward, who took me under her wing as a freshman undergraduate student and coached me through the many challenges of university life. She has always been available to dispense excellent advice and patiently correct my writing, and I'm truly thankful I walked into her office that first semester of college.

I am indebted to all five members of my graduate committee for their hard work and long hours preparing me for this degree. The Weed Research Laboratory staff and students also deserve many thanks for always being willing to help out and offer good advice and ideas. It is not often that you find such a fun and smart group of people to work with and I'm sorry to be leaving them as I finish up this phase of my life.

Finally, I would like to thank my friends and family who gave me encouragement and love every step of the way. They are the source of my strength and help me in more ways than I can count every day of my life.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

Palmer Amaranth.

With its extraordinary growth rate, high seed production, and resistance to multiple herbicidal modes of action, Palmer amaranth (*Amaranthus palmeri* S. Wats.) is recognized as one of the worst weeds to take over cotton, corn, and soybean fields in the United States (Ward et al., 2013). Originating out of the southwestern United States and Mexico (Sauer, 1957), this weed species is capable of growing up to 24 cm tall only four weeks after planting (Sellers et al., 2003) and maintaining a growth rate of 0.21 cm per day to reach final plant heights between 2-3 meters (Horak and Loughin, 2000). The C4 characteristics of Palmer amaranth allow for this rapid growth and biomass accumulation and allow it to survive environments with very high temperatures and low, unpredictable rainfall (Ehleringer, 1983). In one study, Palmer amaranth was found to fix carbon four times faster than cotton (Ehleringer and Hammond, 1987), giving the weed a competitive edge in agricultural conditions and causing significant reductions in crop yields.

Reproductive traits also are weighted in favor of Palmer amaranth. Dioecious in nature, a single female plants can produce between 200-600 thousand seeds in the absence of competition (Keeley, 1987). The small seed size, 1mm in diameter (Keeley et al., 1987) ensures rapid dispersal with seeds surviving in the seedbank more than 40 months (Sosnoskie et al., 2013), especially deeply buried seeds. The male plants are similarly prolific, contributing more than 285 pollen grains per cubic yard of air in some areas

(Walkington, 1959). These pollen grains are highly mobile, moving up to 300m from the pollen source and effectively fertilizing female plants at that distance (Sosnoskie et al, 2012). This dioecious reproductive biology of Palmer amaranth means it is an obligate outcrosser, forcing high levels of genetic diversity (Chandi et al., 2013) which ensure high adaptability.

Even with the above-mentioned weedy traits, Palmer amaranth was not a major problem for American agriculture until the 1990s (Ward et al., 2013). The main reason for this rise in Palmer amaranth control issues and the eventual dubbing of Palmer amaranth as a "superweed" was its development of resistance to a number of herbicidal compounds, most notably the herbicide glyphosate.

Glyphosate Resistance.

Glyphosate is a broad-spectrum herbicide, first sold to farmers by Monsanto in 1974 for post-emergence weed control. It kills plants by binding to the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) protein in the shikimate pathway. Glyphosate competes with the usual substrate, phosphoenolpyruvate (PEP), for binding to the EPSPS-S3P (shikimate-3-phosphate) complex to halt the shikimate pathway, resulting in a lack of aromatic amino acids and a buildup of shikimate, effectively killing the plant. After its commercial introduction, glyphosate rapidly became one of the most widely used herbicides for a number of reasons, including its novel mode of action, low animal and human toxicity, lack of effect on non-target organisms, rapid degradation in the soil, and initial lack of resistance within weedy populations (Duke and Powles, 2008). Despite 20 years of zero resistance and reassurances that resistance to this chemical would not

develop in weeds (Bradshaw et al., 1997), the first instance of glyphosate resistance was reported in 1998 in Australian rigid ryegrass (Powles et al., 1998). Since then, glyphosate resistance has been found in 32 weed species across 25 countries (Heap, 2015).

Glyphosate resistance in Palmer amaranth was first reported in Georgia (Culpepper et al., 2006) and is now present across 25 states (Heap, 2015). For the majority of the Palmer amaranth populations studied, glyphosate resistance is endowed through a gene duplication and amplification mechanism (Sammons and Gaines, 2014). This resistance mechanism was discovered in 2010 (Gaines et al.) and operates via a duplication of the herbicide target gene, 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*), in the genome, resulting in an overproduction of the EPSPS protein. Since no mutation occurs in these *EPSPS* copies, all EPSPS proteins are susceptible to glyphosate binding, but the overproduction ensures that additional EPSPS proteins are available for continued functioning of the shikimate pathway. This is not the only mechanism of glyphosate resistance in Palmer amaranth, with resistance via reduced translocation found in a Mississippi population (Nandula et al., 2012), but it is the only mechanism observed in the populations included in this work.

1.2. PURPOSE AND SCOPE OF THIS WORK

The discovery of gene amplification as a mechanism for herbicide resistance was a novel finding for the field of weed science. Understanding how this resistance trait originated and persists is an important question not only for weed managers seeking to control this weed, but also members of the plant science community interested in the

phenomenon of plant adaptation and evolution. Towards these goals, the research described here focuses on three main areas: glyphosate resistance trait fitness, glyphosate resistance inheritance and overall stability, and the mechanism of *EPSPS* gene duplication.

In the first chapter, dealing with relative fitness, the impact of multiple *EPSPS* copies on Palmer amaranth growth and reproduction was measured by calculating the relative fitness of high *EPSPS* copy number versus single *EPSPS* copy plants across every step of the life cycle. In chapter two, focused on *EPSPS* gene inheritance, the stability of this genic duplication was measured across both meiotic and mitotic cellular divisions. Inheritance of *EPSPS* gene copies was tracked from the parental lines through the F3 generation and within-plant copy number variation was also assessed. Finally, in chapter three, the mechanism of *EPSPS* gene duplication was studied by first assembling the amplicon out as far as possible, then analyzing the assembly for significant features and comparing the consensus with the non-duplicated *EPSPS* allele. In situ hybridization assays were run to determine the genetic location of these duplicated gene copies.

REFERENCES

Anderson KS, Sammons RD, Leo GC, Sikorski JA, Benesi AJ and Johnson KA. (1990) Observation by carbon-13 NMR of the EPSP synthase tetrahedral intermediate bound to the enzyme active site. Biochem, 29:1460-1465.

Bradshaw LD, Padgette SR, Kimball SL, and Wells BH. (1997) Perspectives on glyphosate resistance. Weed Tech, 11: 189-198.

Chandi A, Milla-Lewis SR, Jordan DL, York AC, Burton JD, Zuleta MC, Whitaker JR, and Culpepper, AS. (2013) Use of AFLP markers to asses genetic diversity in Palmer amaranth (*Amaranthus palmeri*) populations from North Carolina and Georgia. Weed Sci, 61:136-145.

Culpepper AS, Grey TL, Vencill WK, Kichler JM, Webster TM, Brown SM, York AC, Davis JW, and Hanna WW. (2006) Glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*) confirmed in Georgia. Weed Sci, 54:620-626.

Duke SO and Powles SB. (2008) Glyphosate: a once-in-a-century herbicide. Pest Man Sci, 64:319-325.

Ehleringer J. (1983) Ecophysiology of *Amaranthus palmeri*, a Sonoran desert summer annual. Oecologia, 57: 107–112.

Ehleringer J and Hammond SD. (1987). Solar tracking and photosynthesis in cotton leaves. Agric For Meteorol, 39: 25–35.

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang JM, Preston C, Leach JE, and Westra P (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proc Natl Acad Sci USA, 107:1029–1034.

Heap, I. (2015) International Survey of Herbicide Resistant Weeds. (http://weedscience.org/summary/moa.aspx). Accessed 20 April 2015.

Horak MJ and Loughin TM. (2000) Growth analysis of four *Amaranthus* species. Weed Sci, 48: 347–355.

Keely PE, Carter CH, and Thullen RJ. (1987) Influence of planting data on growth of Palmer amaranth (*Amaranthus palmeri*). Weed Sci, 35:199-204.

Nandula VK, Reddy KN, Kroger CH, Poston DH, Rimando AM, Duke SO, Bond JA, and Ribeiro DN. (2012) Multiple resistance to glyphosate and pyrithiobac in Palmer amaranth (*Amaranthus palmeri*) from Mississippi and response to flumiclorac. Weed Sci, 60:179-188.

Powles SB, Lorraine-Colwill DF, Dellow JJ and Preston C. (1998) Evolved resistance to glyphosate in rigid ryegrass (*Lolium rigidum*) in Australia. Weed Sci, 46:604-607.

Sammons RD and Gaines TA. (2014) Glyphosate resistance: State of knowledge. Pest Manag Sci, 70: 1367-1377.

Sauer JD. (1957) Recent migration and evolution of the dioecious amaranths. Evol, 11:11-31.

Sellers BA, Smeda RJ, Johnson WG, Kendig JA, Ellersieck, MR. (2003) Comparative growth of six *Amaranthus* species in Missour. Weed Sci, 51:329-333.

Sosnoskie LM, Webster TM, and Culpepper AS. (2013) Glyphosate resistance does not affect Palmer amaranth (*Amaranthus palmeri*) seedbank longevity. Weed Sci, 61:283-288.

Sosnoskie LM, Webster TM, Kichler JM, MacRae AW, Grey TL, and Culpepper AS. (2012) Pollen-mediated dispersal of glyphosate-resistance in Palmer amaranth under field conditions. Weed Sci, 60:366-373.

Walkington DJ. (1959) A survey of the hay fever plants and important atmospheric allergens in the Phoenix, Arizona, metropolitan area. J Allergy, 31:25-41.

Ward SM, Webster TM, and Steckel LE. (2013) Palmer amaranth (*Amaranthus palmeri*): A review. Weed Tech, 27:12-27.

CHAPTER 2. RELATIVE FITNESS OF GLYPHOSATE-RESISTANT PALMER AMARANTH

2.1. INTRODUCTION

Palmer amaranth (*Amaranthus palmeri* S. Wats) is a dicotyledonous weed species native to southern California and now prevalent across most of the southern United States (Steckel 2007). It is highly competitive, capable of growing 2-3 meters in height and producing upwards of 600,000 seeds per female plant when left uncontrolled, resulting in significant yield losses for a number of crop species (Ward et al. 2013).

Adding to the difficulty of Palmer amaranth control is the acquisition of resistance to several herbicides since the late 1980s. Resistance to dinitroaniline herbicides was reported first (Gossett et al. 1992), followed soon after by ALS inhibitors and photosystem II inhibitors (Horak and Peterson 1995; Heap 2012). In 2006, Culpepper et al. confirmed the first appearance of glyphosate resistant Palmer amaranth in Macon County, Georgia. Glyphosate resistant Palmer amaranth has since spread to 13 U.S. states (Heap 2012). Control of glyphosate resistant Palmer amaranth is difficult, especially in cotton fields where the problem first originated. Many cotton farmers are turning to hand-hoeing to rid their fields of Palmer amaranth: between 2000-2005, only 17% of growers hand-hoed their cotton fields, but between 2006-2010, this number rose to 92% of cotton growers (Sosnoskie et al. 2012).

A variety of mechanisms confer resistance to glyphosate in plants, including *EPSPS* (5-enolpyruvylshikimate-3-phosphate synthase) target site point mutations, altered translocation/ vacuolar sequestration, and *EPSPS* gene amplification (Shaner et al. 2012). The Palmer amaranth population from Georgia used in this study is resistant to glyphosate

via *EPSPS* gene amplification (Gaines et al. 2010). In susceptible plants, glyphosate binds to the active site of EPSPS, inhibiting its normal function as a key enzyme in the shikimate pathway, necessary for the production of aromatic amino acids (Steinrucken and Amrhein 1980). When a glyphosate resistant Palmer amaranth plant amplifies the number of *EPSPS* genes, it produces a proportionally increased amount of EPSPS, sufficient to maintain the shikimate pathway in the presence of glyphosate applied at field rates (Gaines et al. 2010). Gene amplification as a mechanism of herbicide resistance has now been reported in two other weed species, *Kochia scoparia* (Wiersma 2012) and *Lolium perenne spp. multiflorum* (Salas et al. 2012). Fitness costs associated with glyphosate resistance have been detected in *Lolium rigidum* (Pederson et al. 2007; Preston and Wakelin 2008) and *Ipomoea purpurea* (Baucom and Mauricio 2004). However, no published studies have examined the relative fitness of Palmer amaranth plants with *EPSPS* amplification-mediated glyphosate resistance compared to glyphosate susceptible plants.

In Palmer amaranth with *EPSPS* gene amplification, mutated copies of the gene have not been found (Gaines et al. 2010). Nevertheless, fitness costs may be associated with this form of glyphosate resistance for other reasons. Some glyphosate resistant Palmer amaranth plants with gene amplification have more than 90 copies of the *EPSPS* gene with a corresponding increase in production of EPSPS protein (Gaines et al. 2010). This level of gene expression could potentially diverts resources from other metabolic processes. In addition, the amplified copies of the *EPSPS* gene are scattered throughout the Palmer amaranth genome, possibly via a mobile genetic element (Gaines et al. 2010). If these additional *EPSPS* gene copies are inserted into functional genes, they may disrupt their activity.

When quantifying the relative fitness of a resistant population versus a susceptible population, a comparison between S and R biotypes will not accurately estimate fitness costs associated with the resistance trait. This comparison assumes that if the resistance allele was removed, the resistant biotype would behave exactly like the susceptible. However, there is a large amount of natural genetic variation in most weed species (Clements et al. 2004), often resulting in a wide spectrum of variation in fitness-related traits, even among individuals collected from the same field population. Differential fitness between R and S biotypes could therefore be due to segregation at fitness-related loci unrelated to the resistance trait.

The best measure of fitness costs associated with resistance would be to determine the change in the frequency of the resistance allele for several generations (Vila-Aiub et al. 2009). This would integrate all the components of fitness across all life history stages. However, tracking allele frequencies for several generations is both time-consuming and expensive, so very few published herbicide resistance fitness studies have attempted this method. Instead, it is common for researchers to use metrics that contribute to and therefore function as surrogates for fitness, such as final plant biomass or total seed production (Vila-Auib et al. 2005; Menalled and Smith 2007; Bagavathiannan et al. 2011). This study uses an F2 population that is segregating for the resistance trait across diverse genetic backgrounds to study the impact of *EPSPS* amplification on fitness.

This research was conducted to compare relative fitness in glyphosate susceptible Palmer amaranth and glyphosate resistant Palmer amaranth with *EPSPS* gene amplification. The objectives were (1) to measure vegetative and reproductive growth parameters in open-pollinating pseudo-F_2 progeny segregating for the resistance trait in a

non-selective controlled greenhouse environment; and (2) to measure the vegetative and reproductive growth parameters of glyphosate resistant and susceptible Palmer amaranth females crossed to the same male under controlled conditions.

2.2. MATERIALS AND METHODS

Plant Material.

Plants grown from Palmer amaranth seed collected from Macon County, Georgia in 2009 were characterized as either glyphosate resistant or susceptible based on a leaf-disc shikimate assay (Shaner et al. 2005) and estimation of *EPSPS* gene copy number via quantitative PCR (details given below). Confirmed glyphosate susceptible and glyphosate resistant plants were used as parents in controlled crosses to develop an F_1 population. No emasculation was necessary because Palmer amaranth is a dioecious species and all crosses were produced in the greenhouse under pollination bags. All parent pairs were comprised of different plants and each pair was bagged separately throughout anthesis.

Levels of glyphosate resistance in individual F_1 progeny were evaluated using a shikimate leaf-disc assay and *EPSPS* copy number was determined using a qPCR-based assay. Confirmed susceptible and resistant F_1 individuals were then crossed to produce the segregating F_2 families used in this study (Table 2.1). These F_2 families were actually "pseudo-F_2s" because Palmer cannot self to produce true F_2 offspring. This study used four F_2 populations segregating for the resistance trait across a range of genetic backgrounds because, as stated earlier, simply comparing a glyphosate susceptible population with a

glyphosate resistant population may detect fitness differences due to other background fitness traits that differ between the two populations.

Fifty pseudo-F_2 seeds from four crosses were plated onto sterile 1% agar in 100 mm by 15 mm petri plates (25 seeds/plate) and placed into a germination chamber set to 35° light/30°C dark (12-hour photoperiod). Once germinating seeds had produced at least 1cm of shoot, they were transferred to 9 x 9 x 9 cm^3 square pots filled with potting soil (Fafard #2 SV, Conrad Fafard Inc., Agawam, MA 01001) and 1 tsp. Osmocote (Osmocote Smart-Release Plant Food, 19-6-12, Scotts Miracle-Gro, Marysville, OH 43041) and placed in a growth chamber set to 30°C, 12-hour days, and 75% humidity. Once plants reached the 2-4 leaf stage, 31 plants were randomly selected from each family, transplanted into 1-L pots with 1 tbsp. Osmocote, and grown under greenhouse conditions. Greenhouse conditions consisted of natural light conditions supplemented with 400W sodium halide lamps to provide a 14-h day length. Daytime temperature was 24°C and nighttime temperature was 18°C. Plants were spaced at least 35cm apart to ensure adequate room for growth and randomized on a weekly basis to average out any position-specific effects. Of the 124 plants, 100 were randomly selected for the primary fitness costs study and the extra 24 were set aside for a reproductive study. This additional study was required because some measurements of growth, including cumulative height and final biomass, could not be measured on plants that were used for controlled crosses due to the growth limitations placed on the plants by enclosing the inflorescences in pollination bags.

Experimental Design.

All 124 F_2 plants were arranged on a greenhouse bench in a completely randomized design, placed 36 cm apart to eliminate competition between plants for light or space. Plants were hand watered daily and re-randomized on the bench weekly. Additional fertilization was supplied twice during the experiment (Miracle-Gro Plant Food Spikes, 2 spikes per pot, Scotts Miracle-Gro, Marysville, OH 43041). The 100 plants used in the main fitness study were left uncovered for the entire period of the experiment, which allowed open pollination to occur. The 24 plants used for the reproductive fitness investigation were individually covered with a micro-perforated pollination bag as soon as the first flowers appeared and then bagged together in sets of three for controlled cross-pollination, as described below.

Estimation of EPSPS Gene Copy Number.

Young leaf tissue was collected from each F_2 plant and genomic DNA was extracted using DNeasy Plant Mini Kits (Qiagen) for estimation of *EPSPS* gene copy number. Concentration and purity of DNA was determined with a ND-1000 Nanodrop spectrophotometer (Thermo Scientific), and only DNA samples with low protein contamination (260/280 ratio of ≥1.8) and low salts/phenol contamination (260/230 ratio of ≥2.0) were used for qPCR.

For the qPCR assay, DNA was diluted to 2 ng/uL in highly purified water. Using PerfeCTa SYBR Green Supermix with ROX (Quanta), and following the supplied protocol, qPCR reactions were set up containing 2.5 uL gDNA template, 1x Perfecta SYBR Green Supermix (containing AccustartTaq DNA Polymerase, dNTPs, SYBR Green I dye, ROX

reference dye, $MgCl_2$, and stabilizers) and 250 nM each of forward and reverse primers for a final reaction volume of 12.5 uL. Primer sequences were identical to those in Gaines et al. (2010).

Primer efficiency curves were created for each primer set using a 1/10x dilution series of genomic DNA from a resistant plant. The *EPSPS* primers (EPSF1: 5'-ATGTTGGACGCTCTCAGAACTCTTGGT-3' and EPSR8: 5'-TGAATTTCCTCCAGCAACGGCAA-3') had an efficiency of 95.2% and the *ALS* primers (ALSF2: 5'-GCTGCTGAAGGCTACGCT-3' and ALFR2: 5'-GCGGGACTGAGTCAAGAAGTG-3') had an efficiency of 95.6%. *ALS* was used as a reference single copy gene because it is known to occur singly at one locus in the genome (Gaines et al. 2010). These efficiencies were very similar and thus directly comparable in subsequent calculations. The genomic DNA templates were run with each primer set in triplicate in 12.5 uL reactions on a 96-well PCR plate. Amplification was performed using the ABI Prism 7000 Real-Time PCR Detection System with the following thermoprofile: 15 minutes at 95°C, 40 cycles of 95°C for 30 seconds and 60°C for 1 minute, and finally a melt-curve analysis to check for primer-dimers. No-template reaction mixes, consisting of 10 uL of Master Mix (1x Perfecta SYBR Green Supermix and 250 nM primers) and 2.5 uL of water, served as the negative controls for this procedure. No primer-dimers and no amplification products were seen in the melt-curve analysis and the controls, respectively.

Threshold cycles (Ct) were calculated by the ABI Prism 7000 program and relative copy number was determined by using a modified version of the $2^{-\Delta\Delta Ct}$ method (Gaines et al., 2010). The *ALS* gene was used a reference gene present in the genome at a copy number of one. *EPSPS* gene copy number was estimated by finding ΔCt = (Ct, *ALS* – Ct, *EPSPS*) and calculating $2^{\Delta Ct}$ to obtain a relative *EPSPS* copy number count.

Plant Growth Measurements.

Starting at the 2-4 leaf stage, plant height was measured weekly as the distance in centimeters from the soil surface to the tip of the apical meristem. Plant volume was also measured weekly by taking two width measurements: one measuring the width of the canopy at its widest point and the other measuring the width at 90 degrees to the first measurement at that same height. Volume was calculated as $\pi r^2 h$, with the average of the two widths equal to r and plant height equal to h. These measurements provide a non-destructive estimate of plant biomass accumulation over the growing season (Bussler et al., 1995). When plants had reached maturity, total aboveground biomass was harvested, separated into vegetative and floral components, and air-dried in large paper bags for two weeks. Vegetative and floral biomass was then weighed and reweighed five days later to confirm all plant tissue was fully dried to constant weight. These measurements were used to calculate total biomass (floral plus vegetative tissue) and harvest index (floral biomass divided by total biomass). Harvest index is normally calculated as the mass of the harvested product divided by the total biomass, but total floral biomass was used as a surrogate measure in this case so both male and female Palmer amaranth could be included.

Measurements of carbon fixation and transpiration rates were taken on a randomly selected subset of 40 plants from the 100 main study plants three times between 10 a.m. and 2 p.m. on July 21, 2011 and repeated on July 22, 2011, when vigorously growing plants were just beginning floral initiation. The youngest fully expanded leaf was removed from the top third of each plant and the distal half of the detached leaf was immediately clamped with the palisade layer upwards in the chamber of a photosynthesis measurement system

(LI-6400XT portable photosynthesis system, LI-COR, Lincoln, NE 68504). System settings were reference CO_2 at 400 ppm, flow rate 400μmol $s^{-1,}$ photosynthetically active radiation 1400 and 10% blue. Each individual measurement was recorded as the mean of five readings taken at three second intervals. Carbon fixation rate for each plant was estimated as the mean of three separate measurements (15 total readings) taken over two days.

Fecundity Measurements.

The 100 pseudo-F2 plants were visually examined twice a week for floral production and the number of days until the first flower emerged was noted. For females, time of first flowering was recorded as the date on which stigmas could first be seen, while for males, first flowering was the date upon which unopened anthers were first visible. Once the anthers had dehisced, pollen was collected from twelve randomly selected male plants, and pollen viability was assessed via a flurochromatic (FCR) assay (Heslop-Harrison and Heslop-Harrision 1970), which uses fluorescein diacetate to check for an intact pollen membrane and the presence of functioning hydrolytic enzymes. Pollen grains with both intact membranes and working enzymes are considered viable and will generate a fluorescence that is detectable under a fluorescence microscope. For this assay, two drops of the florescence/growth mix (1.75*M* sucrose, 3.23m*M* boric acid, 3.05m*M* calcium nitrate, 3.33m*M* magnesium sulfate heptahydrate, 1.98m*M* potassium nitrate, and 7.21m*M* fluoresceindiacetate dissolved in acetone) were pipetted onto a glass slide, and male flowers from the selected plants were gently tapped 3-4 times over the slide to deposit pollen in contact with the liquid. Three separate slides of pollen were collected per plant and immediately examined under a Zeiss Axioskope fluorescence microscope. Using the

blue 450-490 excitation filter, two images of fluorescing pollen grains were captured per slide. Within the viewing field of each image, the number of fluorescing pollen grains and the number of total pollen grains were counted. The total number of fluorescing pollen grains was then divided by the number of total pollen grains to get an average viability ratio per slide, and ratios across all three slides were averaged per plant to obtain mean percent pollen viability.

Additional fecundity-related measurements recorded immediately before total above-ground biomass harvest were length of the longest inflorescence of both males and females and the amount of seed produced by each female. The longest inflorescence length was used as an estimator of both male and female floral production and was measured as the entire length of the apical flowering structure. Seed production was measured by hand threshing seed from each female and weighing the total seed per plant. Palmer amaranth is normally wind pollinated, so conditions in the greenhouse were not optimal for pollen movement. As a result, this measure of seed production may underestimate plant reproductive potential.

Controlled Crossing Fecundity Measurements.

Pollination could not be controlled in the 100 pseudo-F2 plants used for the main fitness study because any controlled crosses would require covering the plants with pollination bags that would restrict growth. Therefore, to estimate reproductive potential of resistant and susceptible plants, controlled crosses were carried out using an additional 24 pseudo-F2 plants as previously described, grown in the same greenhouse environment. Immediately before flowering, each of the 24 plants was bagged with a micro-perforated

pollination bag to prevent pollen movement. Once all plants had started flowering, crosses between two females and one male were assigned according to individual *EPSPS* copy number. Each cross consisted of one high copy number maternal parent and one low copy maternal parent competing for the same pollen. Bags were shaken every other day to promote pollination and groups of pots were randomized across the greenhouse bench on a weekly basis during this period. After pollination had occurred and seeds had matured, each maternal plant was hand-threshed and seed was weighed to determine total seed production. Seed viability tests were carried out by plating 40 seeds from each maternal plant onto 1% agar in 100 mm by 15 mm petri plates and counting the number of seedlings with cotyledons produced after 14 days in a germination chamber set at 35°C light/30°C dark with a 12-hour photoperiod. Seed viability tests were replicated twice.

Male: female Ratio.

Anecdotal evidence has suggested a prevalence of female plants over male plants in Palmer amaranth populations. For this reason, the gender of each plant was recorded and a chi-square analysis was conducted to determine if the population of plants used in this experiment deviated from a 50:50 male: female ratio. A two-sample *t*-test was also conducted to test for any difference in *EPSPS* copy number between males and females, treating gender as the groups and copy number as the independent samples within each group.

Statistical Analysis.

All data analysis was conducted using the JMP statistical package (JMP, SAS Institute Inc., Cary, NC). QQ plots and normality tests were conducted on the growth and reproductive data to check for normal distributions. Final plant height, final plant volume, and longest inflorescence were log transformed to meet the normality and equal variance assumptions of regressions. The initial analyses carried out on this data were regression tests to examine the potential effects of *EPSPS* gene copy number and associated glyphosate resistance on fitness. Each of the growth/fecundity measurements was regressed against individual plant copy number to assess any significant relationships between the two variables.

To differentiate between fitness differences due to *EPSPS* copy number and fitness differences due to family, an analysis of covariance (ANCOVA) was also performed with family as the independent variable, growth and fecundity measurements as the dependent variables, and *EPSPS* copy number as the covariate. The assumptions for ANCOVA (normality, equal variance, independence, equal regression slopes) were met, as tested for by QQ plots and Levene's test. The assumption of equal regression slopes was tested and supported by confirming no interaction between family and *EPSPS* copy number for each dependent variable.

Separate regression analyses between *EPSPS* copy number and growth measures were conducted for each family. The gender data was tested for a 1:1 male : female ratio using a chi-squared analysis for goodness-of-fit.

2.3. RESULTS AND DISCUSSION

Plant Growth and Fecundity Measurements.

There was no statistically significant (α=0.05) relationship between final biomass and *EPSPS* copy number (Figure 2.1 A) for pooled progeny from all four families. There was also no relationship between seed production and *EPSPS* copy number (Figure 2.1 B).

Results from initial pooled-family analysis indicated differences in relationships between *EPSPS* copy number and six of the fitness-related variables. These six variables were seedling plant height, mature plant height and volume, days to first flower, harvest index, and inflorescence length (Figure 2.2). Plant growth related variables (harvest index, height, and volume) showed a positive correlation with *EPSPS* copy number, suggesting a fitness benefit for resistant individuals. The fecundity-related measures of inflorescence length and days to first flower, on the other hand, suggested fitness costs for the resistant plants as a result of later flowering and shorter inflorescences. However, results of ANCOVA indicated that all apparent relationships between fitness-related traits and EPSPS copy number in the pooled data were due to differences between families. That is, any relationships revealed in the initial regressions on pooled individuals from all families were not due to fitness effects associated with *EPSPS* copy number, but to among-family differences in segregation at other fitness-related loci.

This effect was further confirmed when regression analyses were conducted separately on individual families to examine *EPSPS* copy number effects on fitness (Table 2.3). There was a relationship between EPSPS copy number and (1) seedling plant height and days to first flower in family F_2-1, (2) percent seed germination in family F_2-2, and (3)

seed production in family F_2-4, but no relationships were consistent across all families. This is an important point to consider for future fitness studies. In an outcrossing species like Palmer amaranth, high genetic diversity across all loci can be seen (Chandi et al., 2013), making it difficult to pinpoint the true causative sources of fitness differences. Even in the population of plants from this study, where segregating F2 families were used to average out the genetic background, some statistically significant relationships were still observed. Without examining several families, it would be easy to conclude that increased EPSPS copy number directly affected one or more fitness-related traits. However, comparison among different families with different genetic backgrounds showed that no trends were consistent across families and no real relationships were observed.

We can conclude from these results that no fitness costs associated with glyphosate resistance due to EPSPS gene amplification were evident for this Palmer amaranth population when grown under controlled conditions in the greenhouse. As fitness-related measures were taken from plants grown in a greenhouse environment, these results may not directly translate to field conditions.

Controlled Crossing Fecundity Measures.

Results of regression analyses conducted on pollen source determined there was no relationship between *EPSPS* copy number and seed production or seed viability (data not shown). The crossing of both glyphosate resistant and glyphosate susceptible female plants with one male plant ensured exposure to the same pollen source, but no differences in seed production or seed viability between resistant and susceptible mother plants were found. This suggests that there is no maternal reproductive fitness penalty associated with

the *EPSPS* gene amplification mechanism of glyphosate resistance in Palmer amaranth. Relative fitness is defined as the relative ability of different genotypes to pass on their alleles to future generations (Hedrick 2005). Although they are still surrogate measurements for fitness, seed production and germination are direct estimates of the number of potential offspring produced by an individual. The lack of a difference in seed production and seed viability between high and low *EPSPS* copy number female plants in controlled crosses is therefore one of the most notable results indicating the lack of any fitness cost associated with the glyphosate resistance trait.

It is important to note that plants function within a larger ecosystem and the relative fitness of herbicide resistant and susceptible weeds therefore also depends on interactions with other organisms in the agricultural environment. These interactions mean the magnitude of a fitness penalty is dependent on how it behaves in a competitive environment, and these penalties are often more easily measured when competition exists for resources (Reboud and Till-Bottraud 1991). Additionally, it has been documented that fitness costs may vary across an environmental gradient, usually increasing with increased environmental stress such as low nutrient availability (Paris et al. 2008; Raymond et al. 2010). Future studies of fitness costs should test for growth and reproduction differences both in competition studies and across diverse environmental gradients.

Gender Ratio.

A chi-squared test of goodness-of-fit revealed no significant difference in the proportion of males to females, X^2 (1, N=100) = 1.02, p=0.312. There was an observed difference in the emergence time of flowers between the sexes, however, perhaps

contributing to the previous observations suggesting disequilibrium. The t-test for *EPSPS* copy number differences between male and female plants found no difference between the two genders, $t(54) = -0.05$, $p = 0.96$.

Implications.

The lack of fitness penalty associated with the glyphosate resistance mediated by *EPSPS* amplification in Palmer amaranth could have serious repercussions for growers attempting to control this weed. A fitness cost would theoretically favor herbicide susceptible genotypes and shift the weed population back towards increased herbicide susceptibility once herbicide selection was removed (Vila-Aiub et al. 2009). In the case of Palmer amaranth, the lack of fitness costs will likely result in the persistence of glyphosate resistance in the population regardless of selection pressure, leading to the long-term loss of glyphosate as a weed control tool for fields heavily infested with Palmer amaranth.

Results from this study also illustrate the importance of controlling for genetic background when estimating fitness costs, and the need to base fitness estimates on multiple growth, fecundity, and phonological measurements. As this study showed, even full-sib plants from the same family exhibited a high amount of variation in fitness-related measurements, regardless of the level of herbicide resistance. In addition, within-family trends for one fitness-related trait were not duplicated for other traits. These data show that basing estimates of herbicide resistance fitness costs on simple comparisons between resistant and susceptible biotypes for one or two growth-related traits can be highly misleading. Such fitness estimates should be based on multiple growth and fecundity measurements, and must consider the effect of diverse genetic backgrounds.

The results of this study will be also useful in the development of resistance evolution simulation models to predict the trajectory of herbicide resistance and its impact on population dynamics (e.g. Gressel and Segel 1990; Richter et al. 2002). A paper modeling glyphosate resistance in Palmer amaranth demonstrated the potential for this type of tool in testing different weed management strategies (Neve et al. 2010). Including estimates of relative fitness of resistance in future models would lead to more accurate predictions of the impact of different management approaches.

2.4 TABLES

Table 2.1. List of F_2 populations generated and their respective parental 5-enolpyruvylshikimic acid-3-phosphate synthase (*EPSPS*) copy numbers. All *EPSPS* gene number estimates were obtained by quantitative polymerase chain reaction (qPCR), measured against a low-copy internal reference gene (acetolactate synthase, *ALS*).

Generation	Sample size	Maternal *EPSPS* copy number	Paternal *EPSPS* copy number
F2-1	31	1	11
F2-2	31	1	1
F2-3	31	6	39
F2-4	31	47	68

Table 2.2. Significance of 5-enolpyruvylshikimic acid-3-phosphate synthase (*EPSPS*) copy number and family designation on growth and reproductive measurements. An analysis of covariance (ANCOVA) was run on all measurements, using *EPSPS* copy number and family as the two main effects.

	P-value (*EPSPS* copy number)	P-value (family)
Biomass (g)	NS[a]	NS
Harvest index	NS	0.0174*
Average weekly height increase (cm)	NS	NS
Average weekly volume increase (cm^2)	NS	0.0351*
Inflorescence length (cm)	NS	0.0003*
Seedling plant height (cm)	NS	0.0246*
Mature plant volume (cm^2)	NS	0.0019*
Mature plant height (cm)	NS	0.0008*
Days to anthesis	NS	0.0009*
Photosynthesis rate	NS	NS
Transpiration rate	NS	0.0055*
Seed production	NS	0.0243*
Seed germination	NS	NS
Pollen viability	NS	NS

*Significance level set at $\alpha = 0.05$.

[a]Abbreviation: NS, nonsignificant result.

Table 2.3. Effect of 5-enolpyruvylshikimic acid-3-phosphate synthase (*EPSPS*) copy number on growth and reproductive measurements within each family (*N* = 31 per family).

	F_2-1		F_2-2		F_2-3		F_2-4	
	R^2	P	R^2	P	R^2	P	R^2	P
Biomass (g)	0.0209	NS[a]	0.4553	NS	0.0019	NS	0.0018	NS
Harvest index	0.0435	NS	0.0076	NS	0.0201	NS	0.0491	NS
Inflorescence length (cm)	0.0174	NS	0.1288	NS	0.0688	NS	0.0837	NS
Seedling plant height (cm)	0.3117	0.0022*	0.0042	NS	0.0220	NS	0.0070	NS
Mature plant volume (cm^2)	0.0077	NS	0.0095	NS	0.0573	NS	0.0435	NS
Mature plant height (cm)	0.0238	NS	0.0009	NS	0.1539	NS	0.0469	NS
Days to anthesis	0.1534	0.0301*	0.0452	NS	0.0137	NS	0.0199	NS
Photosynthesis rate	0.0000	NS	0.0402	NS	0.1287	NS	0.0176	NS
Transpiration rate	0.0000	NS	0.1703	NS	0.0104	NS	0.0094	NS
Seed production	0.1090	NS	0.0426	NS	0.0005	NS	0.3673	0.0367*
Seed germination	0.0641	NS	0.2407	0.0387*	0.0753	NS	0.0374	NS
Pollen viability	0.0000	NS	0.6452	NS	0.5037	NS	0.0275	NS

*Significance level set at $\alpha = 0.05$.

[a]Abbreviation: NS, nonsignificant result.

2.5 FIGURES

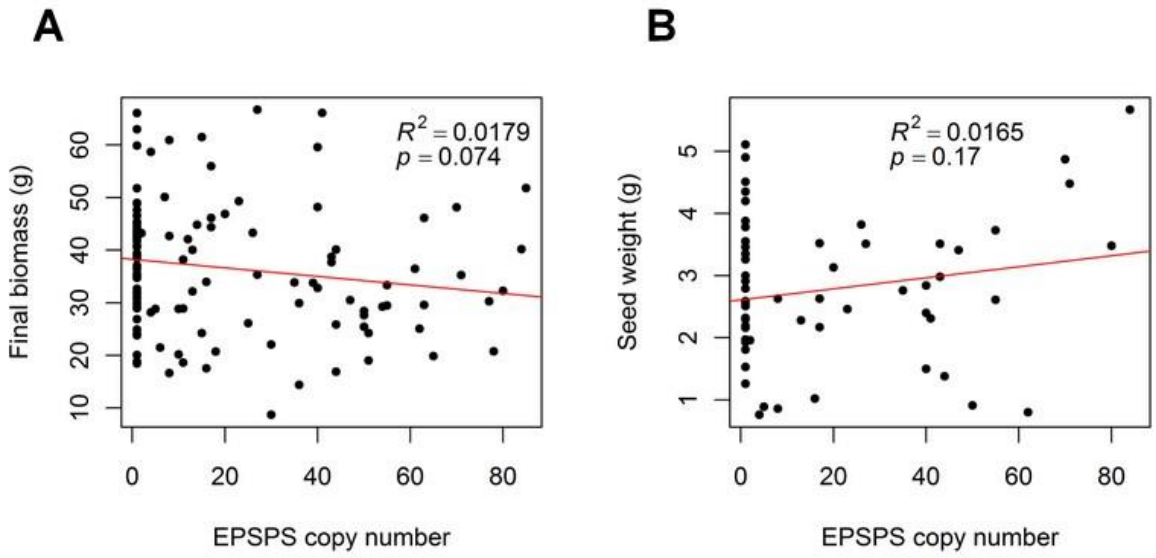


Figure 2.1. (A) Regression of final plant biomass (g) on *EPSPS* (5-enolpyruvylshikimic acid-3-phosphate synthase) copy number. (B) Regression of total seed weight per plant (g) on *EPSPS* copy number.

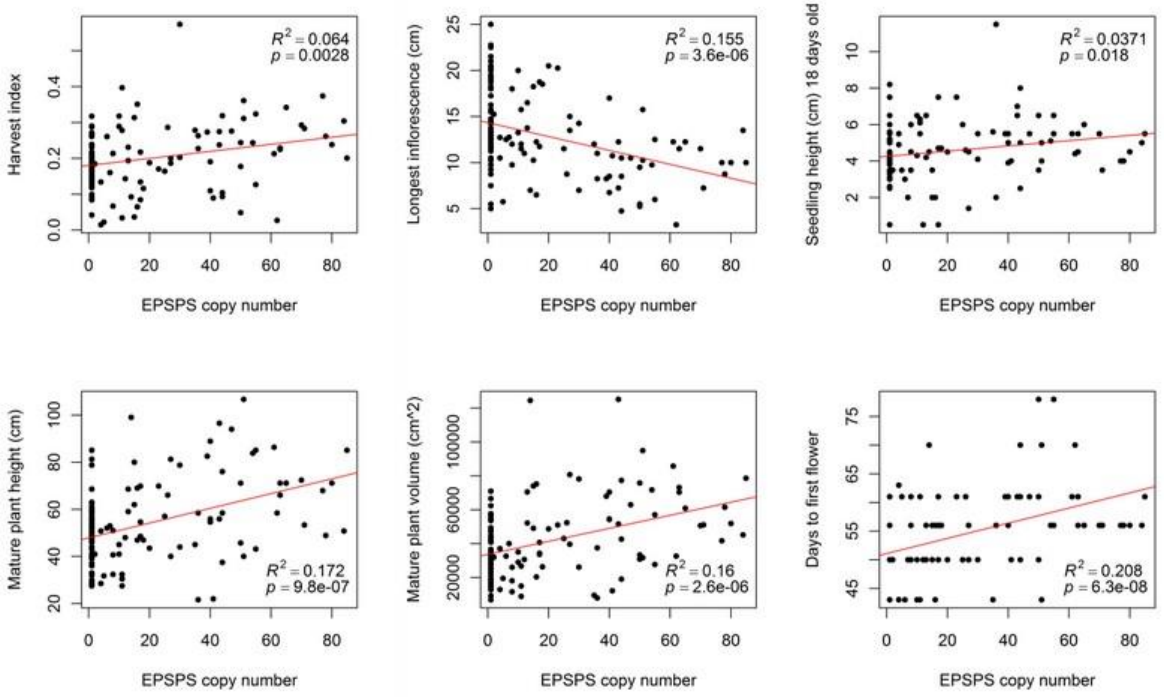


Figure 2.2. Regression of a subset of the fitness measures on *EPSPS* (5-enolpyruvylshikimic acid-3-phosphate synthase) copy number.

REFERENCES

Bagavathiannan MV, Norsworthy JK, Jha P, and Smith KL. (2011) Does Resistance to Propanil or Clomazone Alter the Growth and Competitive Abilities of Barnyardgrass (*Echinochloa crus-galli*). Weed Sci. 59:353-358.

Baucom RS and Mauricio R. (2004) Fitness costs and benefits of novel herbicide tolerance in a noxious weed. Proc. Natl. Acad. Sci. U.S.A. 101:13386-13390.

Bussler BH, Maxwell BD, and Puettmann KJ. (1995) Using plant volume to quantify interference in corn (*Zea mays*) neighborhoods. Weed Sci. 43:586-594.

Chandi A, Milla-Lewis SR, Jordan DL, York AC, Burton JD, Zuleta MC, Whitaker JR, and Culpepper AS. (2013) Use of AFLP Markers to Assess Genetic Diversity in Palmer Amaranth (*Amaranthus palmeri*) Populations from North Carolina and Georgia. Weed Sci. 61:136-145.

Clements DR, DiTommaso A, Jordan N, Booth BD, Cardina J, Doohan D, Mohler CL, Murphy SD, and Swanton CJ. (2004) Adaptability of plants invading North American cropland. Agric. Ecosyst. Environ. 104:379-398.

Culpepper AS, Grey TL, Vencill WK, Kichler JM, Webster TM, Brown SM, York AC, Davis JW, and Hanna WW. (2006) Glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*) confirmed in Georgia. Weed Sci. 54:620-626.

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang JM, Preston C, Leach JE, and Westra P. (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proc Natl Acad Sci USA. 107:1029–1034.

Gossett BJ, Murdock EC, and Toler JE. (1992) Resistance of Palmer amaranth (*Amaranthus palmeri*) to the dinitroaniline herbicides. Weed Technol. 6:587-591.

Gressel J and Segel LA. (1990) Modeling the effectiveness of herbicide resistance and mixtures as strategies to delay or preclude resistance. Weed Technol. 4:186-198.

Heap I. (2012) The International Survey of Herbicide Resistant Weeds. Available at www.weedscience.com. Accessed December 29, 2012.

Hedrick PW. (2005) Genetics of populations. 3rd ed. Sudbury, MA: Jones and Bartlett Inc. 736 pp.

Heslop-Harrison J and Heslop-Harrison Y. (1970) Evaluation of pollen viability by enzymatically induced fluorescence: Intracellular hydrolysis of fluorescein diacetate. Stain Tech. 45:115-120.

Horak MJ and Peterson DE. (1995) Biotypes of Palmer amaranth (*Amaranthus palmeri*) and common waterhemp (*Amaranthus rudis*) are resistant to imazethapyr and thifensulfuron. Weed Technol. 9:192-195.

Menalled FD and Smith RG. (2007.) Competitiveness of herbicide-resistant and herbicide-susceptible kochia (*Kochia scoparia* [L.] Schrad.) under contrasting management practices. Weed Biol. Manag. 7:115-119.

Neve P, Norsworthy JK, Smith KL, and Zelaya IA. (2010.) Modeling evolution and management of glyphosate resistance in *Amaranthus palmeri*. Weed Res. 51: 99-112.

Paris M, Roux F, Berard A, and Reboud X. (2008) The effects of the genetic background on herbicide resistance fitness cost and its associated dominance in *Arabidopsis thaliana*. Heredity 101:499-506.

Pederson BP, Neve P, Andreasen C, and Powles SB. (2007) Ecological fitness of a glyphosate-resistant *Lolium rigidum* population: growth and seed production along a competition gradient. Basic Appl. Ecol. 8:258-268.

Preston C and Wakelin AM. (2008) Resistance to glyphosate from altered herbicide translocation patterns. Pest Manag. Sci. 64:372-376.

Raymond B, Wright DJ, and Bonsall MB. (2010) Effects of host plant and genetic background on the fitness costs of resistance to *Bacillus thuringiensis*. Heredity 106:281-288.

Reboud X and Till-Bottraud I. (1991) The cost of herbicide resistance measured by a competition experiment. Theor. Appl. Genet. 82:690-696.

Richter O, Zwerger P, and Bottcher U. (2002) Modeling spatio-temporal dynamics of herbicide resistance. Weed Res. 42:52–64.

Salas RA, Dayan FE, Pan Z, Watson SB, Dickson JW, Scott RC, and Burgos NR. (2012) *EPSPS* gene amplification in glyphosate-resistant Italian ryegrass (*Lolium perenne* ssp. *multiflorum*) from Arkansas. Pest Manag. Sci. 68:1223-1230.

Shaner D, Nadler-Hassar T, Henry WB, and Kroger C. (2005) A rapid *in vivo* shikimate accumulation assay with excised leaf discs. Weed Sci. 53:769-774.

Shaner DL, Lindenmeyer RB, and Ostlie MH. (2012) What have the mechanisms of resistance to glyphosate taught us? Pest Manag. Sci. 68:3-9.

Sosnoskie LM, Culpepper AS, Grey TL, and Webster TM. (2012) Compensatory growth in Palmer amaranth: effects on weed seed production and crop yield. Page 99 in 2012

Proceedings of the Western Society of Weed Science. Reno, NV: Western Society of Weed Science.

Steckel, L. E. (2007) The dioecious *Amaranthus* spp.: Here to stay. Weed Tech. 21:567-570.

Steinrucken HC and Amrhein N. (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochem. Biophys. Res. Commun. 94:1207-1212.

Vila-Aiub MM, Neve P, Steadman KJ, and Powles SB. (2005) Ecological fitness of a multiple herbicide-resistant *Lolium rigidum* population: dynamics of seed germination and seedling emergence of resistant and susceptible phenotypes. J. Appl. Ecol. 42:288-298.

Vila-Aiub MM, Neve P, and Powles SB. (2009) Fitness costs associated with evolved herbicide resistance alleles in plants. New Phytol. 184:751-767.

Ward SM, Webster TM, and Steckel LE. (2013) Palmer amaranth (*Amaranthus palmeri*): A review. Weed Tech. 27:12—27.

Wiersma A. (2012) Regional whole plant and molecular response of *Kochia scoparia* to glyphosate. M.S. thesis. Fort Collins, CO: Colorado State University. 64p.

CHAPTER 3. MITOTIC AND MEIOTIC INSTABILITY OF DUPLICATED *EPSPS* GENES

3.1 INTRODUCTION

Understanding the genetic basis and mode of inheritance of herbicide resistance is essential for predicting its persistence, evolutionary trajectory and rate of spread in a given weed species (Jasieniuk et al. 1996; Neve 2007). Such information is especially valuable for resistance to glyphosate, currently the most widely used herbicide globally (Beckie 2011). Reported incidences of glyphosate resistance have increased steadily since the first case was reported in Australia in 1995 (Pratley et al. 1999) and it has now been confirmed in 24 weed species worldwide (Heap 2013). Most glyphosate resistance mechanisms investigated to date are inherited as a dominant or semidominant allele at a single locus (Christoffers and Varanasi 2010). Examples include glyphosate resistance based on vacuolar sequestration in horseweed (*Conyza canadensis* L. Cronq.) (Zelaya et al. 2004; Ge et al. 2010); resistance resulting from reduced herbicide translocation in rigid ryegrass (*Lolium rigidum* Gaud.) (Wakelin and Preston 2006; Preston and Wakelin 2008) and resistance associated with amino acid substitutions at the proline$_{106}$ position within the EPSPS gene in several weed species, including goosegrass (*Eleusine indica* (L.) Gaertn.) (Ng et al. 2004) and Italian ryegrass (*Lolium perenne* (L). ssp. *multiflorum* Lam.) (Jasieniuk et al. 2008). Simarmata et al. (2005) reported glyphosate resistance in a Californian rigid ryegrass population was inherited as two semi-dominant independently segregating alleles. However, Yu et al. (2007) described a South African population of rigid ryegrass with resistance to multiple herbicides including glyphosate; these plants had accumulated

resistance alleles at different loci, including combining the proline$_{106}$ EPSPS target site mutation with reduced translocation of glyphosate. As already described, these resistance mechanisms can be inherited independently as single alleles. It is possible, therefore, that glyphosate resistance apparently controlled by a two-locus system may in fact consist of two different mechanisms, each controlled independently at a single locus.

An exception to resistance mechanisms mediated at one or two loci is glyphosate resistance based on amplification of the EPSPS gene. This increase in gene copy number results in overexpression of the EPSPS enzyme targeted by glyphosate, enabling resistant plants to produce sufficient enzyme to maintain the shikimate pathway even in the presence of glyphosate (Gaines et al. 2010; Powles 2010). Glyphosate resistance linked to EPSPS gene amplification was first confirmed in a Palmer amaranth (*Amaranthus palmeri*) population from Georgia (Gaines et al. 2010); since then, additional examples of this glyphosate resistance mechanism have been found in Palmer amaranth populations from North Carolina (Chandi et al. 2012), Mississippi (Ribeiro et al. 2011), and New Mexico (Mohseni-Moghadam et al. 2013). EPSPS gene amplification has also been reported in glyphosate resistant populations of kochia (Wiersma et al. 2015) and Italian ryegrass (Salas et al. 2012).

Investigations to date of the inheritance of glyphosate resistance associated with EPSPS gene amplification in Palmer amaranth have shown inconclusive results. Mohseni-Moghadam et al. (2013) did not find Mendelian segregation ratios consistent with a single gene mechanism among resistant (R) and susceptible (S) phenotypes in F1 and pseudo-F2 families. Chandi et al. (2012) reported apparent single-gene segregation between R and S phenotypes in some BC1F1 families derived from crosses between R and S parents;

however, these authors also found that segregation in other BC1F1 families did not conform to a single-gene model, and that results for some families also potentially conformed to a two-gene additive model. Both Mohseni-Moghadam et al. (2013) and Chandi et al. (2012) reported no differences in progeny from reciprocal R x S and S x R crosses, and concluded this form of glyphosate resistance is under nuclear control with no maternal effects. Neither of these studies quantified EPSPS gene copy number in the F1 or pseudo-F2 progeny: the authors estimated R : S segregation ratios by exposing progeny families to screening doses of glyphosate and recording the proportion of survivors. A major limitation of this approach is that the genomes of Palmer amaranth plants capable of surviving field rates of glyphosate may contain from 10 to over 100 EPSPS gene copies (Gaines et al. 2011); assigning these diverse genotypes to a single R phenotype fails to detect this potentially important variation in progeny gene copy number. Gaines et al. (2010; 2011) reported EPSPS copy numbers ranging from 1 to over 100 in a pseudo-F2 family of 54 plants; the gene copy number of one of these F2 plants was greater than the combined copy numbers of the F1 parents, indicating amplified EPSPS gene transmission via sexual reproduction is not merely additive, but more complex and potentially unstable. The objective of the research reported here, therefore, was to further investigate and characterize the inheritance and intergenerational stability of amplified *EPSPS* gene copies in Palmer amaranth for both sexual and asexual reproduction.

3.2. MATERIALS AND METHODS

Plant Material and EPSPS Copy Number Estimation.

Parent plants were grown from seed collected in 2009 from a resistant Palmer amaranth population in Macon County, Georgia. Seeds were plated onto sterile 1% agar in 100 mm by 15 mm petri plates (25 seeds/plate) and placed in a germination chamber set to 35 C light/30 C dark with a 12-hour photoperiod. Germinated seeds with at least 1 cm of visible shoot were transferred into potting soil (Fafard #2 SV, Conrad Fafard Inc., Agawam, MA 01001) in 9 cm^3 plastic pots in a growth chamber set at 30 C, 75% humidity and a 12-hour photoperiod. One teaspoon of slow-release fertilizer (Osmocote Smart-Release Plant Food, 19-6-12, Scotts Miracle-Gro, Marysville, OH 43041) was added to the soil in each pot at planting. Plants were transplanted at the 2 to 4 leaf stage into 1-L pots and transferred to a Colorado State University greenhouse where they were grown under 400-W sodium halide lamps to provide a 14-hour photoperiod and temperatures of 24 C day/18 C night. All pots were spaced at least 35 cm apart on a greenhouse bench and re-randomized weekly.

Plants used as parents in controlled crosses were initially confirmed as glyphosate resistant or susceptible using the leaf-disc shikimate assay described by Shaner et al. (2005). Young leaf tissue was collected from individual parent plants and total genomic DNA was extracted using DNeasy Plant Mini Kits (Qiagen, Valencia, CA 91355) for estimation of *EPSPS* copy number via qPCR as described by Giacomini et al. (2013). The acetolactate synthase (*ALS*) gene was used as a reference because it occurs as a single copy at a known single locus within the Palmer amaranth genome (Gaines et al. 2010). Relative

to this ALS standard, all plants classified as susceptible based on the shikimate assay had one relative *EPSPS* gene copy and all plants classified as resistant had 10 or more *EPSPS* gene copies. In the rest of this paper, when a plant is labeled as "single copy *EPSPS*", it means that plant had one *EPSPS* gene copy relative to the *ALS* gene.

Generation and Testing of F1 and Pseudo-F2 Families.

Controlled crosses were carried out by enclosing one male and one female plant together in a micro-perforated pollination bag as soon as the first flowers appeared and before anthesis or stigma exsertion. Bags were shaken daily to ensure pollen transfer to the stigmas. Ten F1 families were produced, each from a controlled cross between a separate pair of parents. Details of the parent pairs are given in Table 3.1. F1 plants were germinated and grown, and EPSPS copy numbers were determined as previously described. Gene copy number estimates were obtained for 6 to 18 progeny from each F1 family. Shikimate detection using the in vitro leaf disk shikimate assay (Shaner et al. 2005) was also performed on all parental and F1 plants to test for correlation of *EPSPS* copy number with glyphosate resistance.

F1 plants were crossed by enclosing one female plant and one male plant from the same F1 family together in a micro-perforated pollination bag, as previously described. Palmer amaranth is dioecious, so F1 plants cannot be self–pollinated, and full-sib F1 pairs must be crossed to generate a pseudo-F2 generation; these progeny will be referred to hereafter as F2 families. Four F2 families representing all combinations of crosses (R x R, S x S, and R x S), were selected for *EPSPS* copy number examination via qPCR. Thirty-one

plants from each family were grown under greenhouse conditions as previously described until they were large enough for sampling of young leaf tissue for DNA extraction.

To test for glyphosate resistance prevalence in each generation, a spray test was conducted on F1, F2, and F3 seeds. The F3 seeds were collected from controlled crosses between single female-male pairs of full-sib F2 plants and these seeds were grown under greenhouse conditions alongside seeds from the F1 and F2 populations. At the 3-4 leaf stage, 15 plants from all F1, F2, and F3 populations were sprayed with 840 g ae/ha glyphosate (Roundup Weathermax®, 540 g ai L^{-1}, Monsanto, 800 N Lindbergh Blvd., St. Louis, MO 63167) with 1% ammonium sulfate. Plant mortality was recorded 14 days after treatment (DAT).

Generation and Testing of Clonally Propagated Shoots.

Starting with a high *EPSPS* copy number female plant (n = 122), clones were produced by excising side shoots that were at least 7 cm in length. To reduce transpiration losses, shoots were stripped of all lower leaves until just 3-4 leaves remained. The basal ends of the shoots were dipped in rooting hormone (Root Boost, 0.1% indole-3 butyric acid, GardenTech, Palatine, IL 60095) and planted into potting soil (Fafard #2 SV, Conrad Fafard Inc., Agawam, MA 01001). Potted clones were placed under a humidity dome and allowed to develop roots in the greenhouse under conditions of 24°C/18°C day/night temperatures and a 14-h day length.

After ten generations of cloning, in which each generation represented the creation of a single side shoot clone from the previous generation plant, 12 plants were randomly selected for analysis. *EPSPS* copy number and glyphosate response were measured using

the previously described quantitative PCR method and leaf disc shikimate assay. All plants were clones derived from the same original plant with 122 *EPSPS* gene copies. Additional shoots were cloned directly from these 12 plants and assayed for *EPSPS* copy number to examine any differences that occur within a single generation of cloning.

To investigate whether somaclonal variation in EPSPS copy number occurs within an individual plant, a second study was conducted on a different glyphosate resistant Palmer amaranth population, originating from the same Georgia county as the population used in the previous cloning study. Five plants from this population were analyzed, using qPCR to determine *EPSPS* gene copy number from six lateral meristems of each plant. Unlike the previous experiment, these side shoots were not excised and cloned. Rather, tissue was collected from the youngest leaves of each shoot while still connected to the plant.

Statistical Methods.

All data was analyzed using R v3.0.1 (http://www.r-project.org/). Shapiro-Wilk normality tests were carried out on all F1 and pseudo-F2 progeny to test for non-normal distributions of offspring *EPSPS* copy number. This type of analysis identified any families with skewed, bimodal, and/or non-symmetric distributions. A linear regression analysis to test for trans-generational *EPSPS* copy number stability was run between cumulative parental *EPSPS* gene copy number and mean offspring *EPSPS* gene copy number (confirmed to be normally distributed via Shapiro-Wilk test) across F1 and F2 populations. A lack -of-fit test was performed on the data to ensure a linear regression model was appropriate by following the procedures laid out by Kniss and Streibig (2015). The lack-of-

fit test confirmed a linear model fit this data. One-way ANOVAs (analysis of variance) with Tukey honest significant differences (HSD) post hoc tests were run on all clonal and within-plant copy number data to test for significant copy number differences. Values of $p < 0.05$ were considered to be statistically significant for all tests.

3.3. RESULTS AND DISCUSSION

Non-Mendelian Inheritance of Increased EPSPS Copy Number.

Similar to the results reported by Chandi et al. (2012) and Mohseni-Moghadam et al. (2013), the F1 and F2 families examined here did not show Mendelian segregation for resistance when individuals were pooled into R and S phenotypic classes based on shikimate assays. More detailed examination of individual F1 and F2 offspring revealed wide within-family ranges of EPSPS gene copy number that - unlike a classical quantitative trait - did not consistently follow a normal distribution in all families (Figure 3.1 and Table 3.2). Linear regression of mean offspring EPSPS copy number on combined parental copy number showed a significant relationship between the two variables (see Figure 3.2): the offspring of two high copy number parents typically also have higher copy numbers. However, a similar regression analysis of mean offspring copy numbers against maternal or paternal copy numbers was not significant, indicating neither maternal nor paternal influence on offspring copy number. This more detailed analysis confirms previous observations that glyphosate resistance in Palmer amaranth does not appear to be maternally inherited or subject to maternal effects (Chandi et al. 2012).

Transgressive segregation occurred in five out of 11 F1 families and in all four F2 families, with some offspring having EPSPS copy numbers that either exceeded the high copy number parent or were less than the low copy number parent. It appears, therefore, that the mechanism for EPSPS gene amplification in Palmer amaranth is capable of generating rapid gains and/or losses of copy number in a single cycle of sexual reproduction. Remarkably, crossing two single-copy susceptible F1 plants produced an F2 family consisting of 29 single-copy plants and two individuals with amplified EPSPS copy numbers of 12 and 23 respectively (see family F2-3 in Figure 3.1). This indicates that even in the absence of glyphosate, EPSPS amplification can occur within a single generation in the offspring of susceptible single-copy plants. This is a result with alarming implications for managing glyphosate resistance in Palmer amaranth, especially given the persistent occurrence of high copy individuals in our F1 and F2 families without glyphosate selection, and the lack of fitness costs associated with EPSPS amplification reported previously (Giacomini et al. 2014; Vila-Aiub et al. 2014).

Mitotic instability of increased EPSPS copy number.

In the clonal lines originating from a high *EPSPS* copy number (n = 122) female Palmer amaranth plant, ten generations of clonal propagation via shoot cuttings resulted in plants with *EPSPS* copy numbers ranging from 40-180 (Figure 3.3). As these cloned plants had never gone through a sexual reproductive phase, any change in gene copy number must have occurred in mitotically dividing cells. This same phenomenon was also observed in plants in the next (11^{th}) generation of clones (Table 3.3), confirming changes in *EPSPS* copy number after only a single generation of cloning. This suggests that an active

and unpredictable mechanism drives gains and losses of *EPSPS* gene copies, and that such changes can occur during mitosis as well as meiosis.

Further evidence for active mitotic changes in EPSPS copy number was provided by examining lateral meristems within individual plants. All five plants tested showed significant differences in *EPSPS* copy number among the six lateral meristems sampled (Table 3.4), and Tukey honest significant differences (HSDs) were found between many of the side shoots within each plant (Figure 3.4). The largest absolute difference in *EPSPS* copy number within a single plant ranged from 7 to 30, with plants exhibiting the highest overall mean *EPSPS* copy numbers also producing the widest ranges. Lateral branches were sampled from top to bottom of each plant, but no significant correlation between copy number and lateral meristem position was found ($r = 0.09$, $n = 30$, $p = 0.10$). There was also no difference between male and female plants in the occurrence and extent of this internal copy number variation (data not shown).

The somatic mosaicism observed in our shoot sampling study provides one possible explanation for the wide ranges in EPSPS copy number and the transgressive segregation that we observed in the F1 and F2 families. Male and female Palmer amaranth plants flower not only at the apical meristem but also at the side branches, so they could produce pollen and ovules with different copy numbers depending on the extent of EPSPS amplification within and among lateral meristems. Unequal meiotic recombination may also contribute to varying copy numbers within gametes, including loss of gene copies. It should also be noted that the EPSPS copy number for plants used as parents in our controlled crosses was determined by extracting genomic DNA from a single leaf, so a copy number estimate obtained via qPCR may not be representative of the whole plant. In addition, seeds were

collected from all branches and pooled across the maternal plant. Maternal and paternal parents in a controlled cross with different *EPSPS* copy numbers among the flowering branches could be a major contributing factor to the wide variation and unpredictable segregation for copy number that we observed in the F1 and F2 families.

Somatic mosaicism could also explain the unexpected appearance of higher copy number offspring from two presumed single-copy plants, if such plants had additional gene copies in the unsampled lateral meristems. However, all plants that we designated as single-copy based on qPCR analysis were susceptible to glyphosate, and a consistent association between the glyphosate susceptible phenotype and a presumed single copy genotype based on sampling one leaf has been reported in other studies (Gaines et al. 2010, 2011; Giacomini et al. 2014; Vila-Aiub et al. 2014). It is possible that low levels of somatic mosaicism occur in susceptible plants, with varying numbers of EPSPS copies in some lateral meristems but insufficient levels of gene duplication to confer glyphosate resistance at the screening rates used. This deserves further investigation.

The mechanism of EPSPS gene amplification in Palmer amaranth is still unknown. However, the somatic mosaicism for copy number, and the rapid gains and losses of gene copies that we observed in both sexually produced and vegetatively cloned generations, could be the result of unequal mitotic and/or meiotic recombination. Mitotic recombination producing novel phenotypes through genomic rearrangement, including sequence gains and losses, has been shown to occur in angiosperms (Wicker et al. 2007; Yuan et al. 2011). This phenomenon should be considered as a potential contributor to the unpredictable and unstable transmission of increased EPSPS gene copy numbers seen in

this study within individual plants, among F1 and F2 sibs within families, and from generation to generation whether cloned or sexually produced.

Conclusion.

Inheritance of glyphosate resistance via increased *EPSPS* gene copy number is a complex process. This work shows clear evidence that copy number is inherited in a non-Mendelian fashion, resulting in a high amount of variation in *EPSPS* gene copy number among the offspring of all crosses. This variation is further augmented by the mitotic variation in copy number occurring within individual plants, producing a somatic mosaic phenotype for all Palmer amaranth plants assayed. The results of all of this variation include (1) progeny with higher *EPSPS* gene copy number than the high copy number parent (transgressive segregation) and (2) rapid loss and gain of *EPSPS* gene copies, allowing for a very quick response time to selection. Once again, Palmer amaranth is proving itself to be a highly adaptable weed species.

The presence of these additional *EPSPS* gene copies despite a lack of glyphosate applications to act as a positive selection pressure indicates that this resistance trait will likely persist in natural populations for many years. Of even greater concern is the potential for the duplicated *EPSPS* gene copies to allow for higher than average mutation rates as a result of genetic redundancy. Mutations in one or a few of these gene copies will have little impact on the overall fitness of the plant because the rest of the gene copies will compensate for any deleterious mutations. With a higher chance of hitting upon a target site mutation conferring glyphosate resistance (e.g. Pro106Ser, Thr102Ile, etc.), the problem of glyphosate resistance may be compounded by additive resistance mechanisms.

Further work is needed to see if duplicated *EPSPS* genes display higher mutation rates. Additional studies are also needed to understand how Palmer amaranth is able to gain and lose gene copies so quickly.

3.4. TABLES

Table 3.1. Maternal and paternal *EPSPS* gene copy numbers of F1 and pseudo-F2 crosses.

Generation	N	Maternal *EPSPS* copy number	Paternal *EPSPS* copy number
F1-1	10	1	42
F1-2	16	1	68
F1-3	18	1	28
F1-5	16	80	1
F1-7	6	66	1
F1-8	7	70	1
F1-9	17	108	1
F1-10	6	75	1
F1-12	9	61	12
F1-16	8	123	59
F1-19	10	1	70
F2-2	31	2	13
F2-3	31	1	1
F2-8	31	7	52
F2-9	31	47	68

Table 3.2. Shapiro-Wilk tests for normality on each F1 and pseudo-F2 family. P-values less than α=0.05 are reported and indicate non-normal distributions of progeny *EPSPS* gene copy number.

Family	W¥	P value
F1-1	0.7412	0.00279
F1-2	0.9608	NS
F1-3	0.2527	1.1E-08
F1-5	0.921	NS
F1-7	0.9601	NS
F1-8	0.986	NS
F1-9	0.9446	NS
F1-10	0.6663	0.00268
F1-12	0.8609	NS
F1-16	0.9842	NS
F1-19	0.936	NS
F2-2	0.7464	6.07E-06
F2-3	0.2677	2.48E-11
F2-8	0.8025	5.75E-05
F2-9	0.9783	NS

¥ Shapiro-Wilk test statistic

NS = non-significant

Table 3.3. *EPSPS* gene copy numbers of Palmer amaranth clones after ten cycles (generations) of cloning and from the next cloning cycle (11th generation).

Plant ID	10th generation *EPSPS* copy number	11th generation *EPSPS* copy numbers
A-1	40	35; 41
A-2	63	50; 81
B-1	117	90; 146
B-2	149	176; 184
C-1	128	90; 125; 145; 148
C-2	114	122; 127; 144; 168

Table 3.4. ANOVA results of side shoot *EPSPS* copy numbers from each glyphosate resistant Palmer amaranth plant (R1-R5). Significant p-values indicate statistically different *EPSPS* copy numbers between branches from the same plant.

Plant ID	df	SS	MS	F value	p value
R1	5	2088.0	417.5	313.1	2.81e-12***
R2	5	99.6	19.9	25.6	5.17e-06***
R3	5	190.9	38.2	14.3	1.05e-04***
R4	5	819.6	163.9	18.2	3.12e-05***
R5	5	1380.4	276.1	33.6	1.18e-06***

Significance codes: <0.01 '***' 0.01 '**' 0.05 '*'

3.5. FIGURES

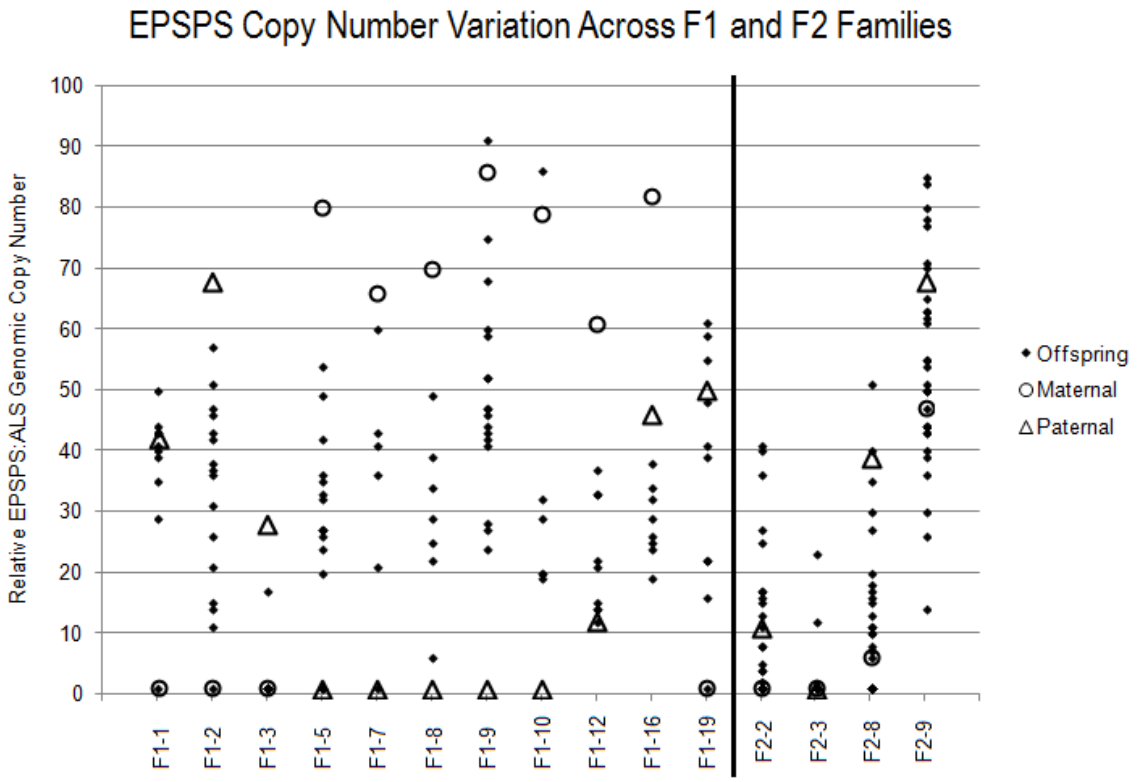


Figure 3.1. Parental and offspring *EPSPS* gene copy numbers for F1 and pseudo-F2 crosses. Each column represents a single cross or family, consisting a maternal plant (○), a paternal plant (△), and the progeny (◆).

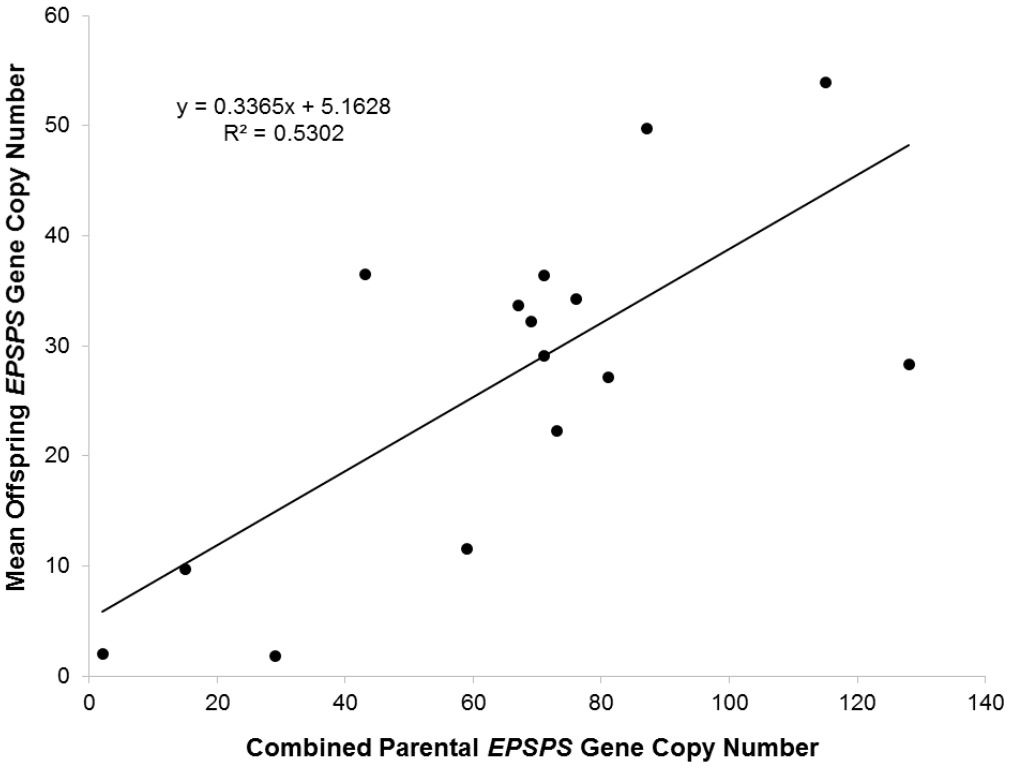


Figure 3.2. Linear regression of additive parental *EPSPS* gene copy number (maternal plus paternal) against mean offspring *EPSPS* gene copy number. N = 15, R^2 = 0.5303, p-value = 0.002.

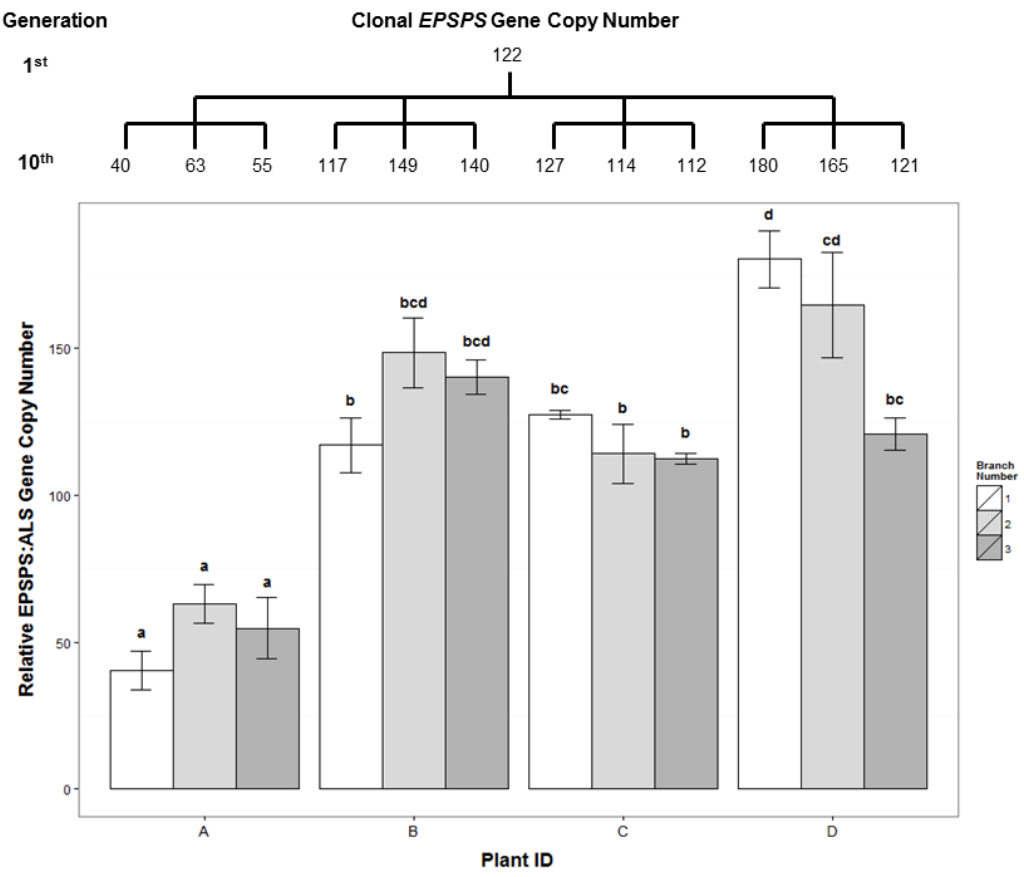


Figure 3.3. Variation in *EPSPS* gene copy number after ten cycles of cloning from a high *EPSPS* copy number plant (original plant copy number = 122). ANOVA, df = 11, F value = 22.27, p value = 5.92e-10. Mean ± SE for each clone with Tukey HSD test results shown as letters above columns. Means not sharing the same letter are significantly different. Tree above graph shows mean number of *EPSPS* gene copies for each clone and relationship between clones. The tenth generation of clones were taken from four plants (A-D), with each plant producing three clones, indicated by the branches of the tree.

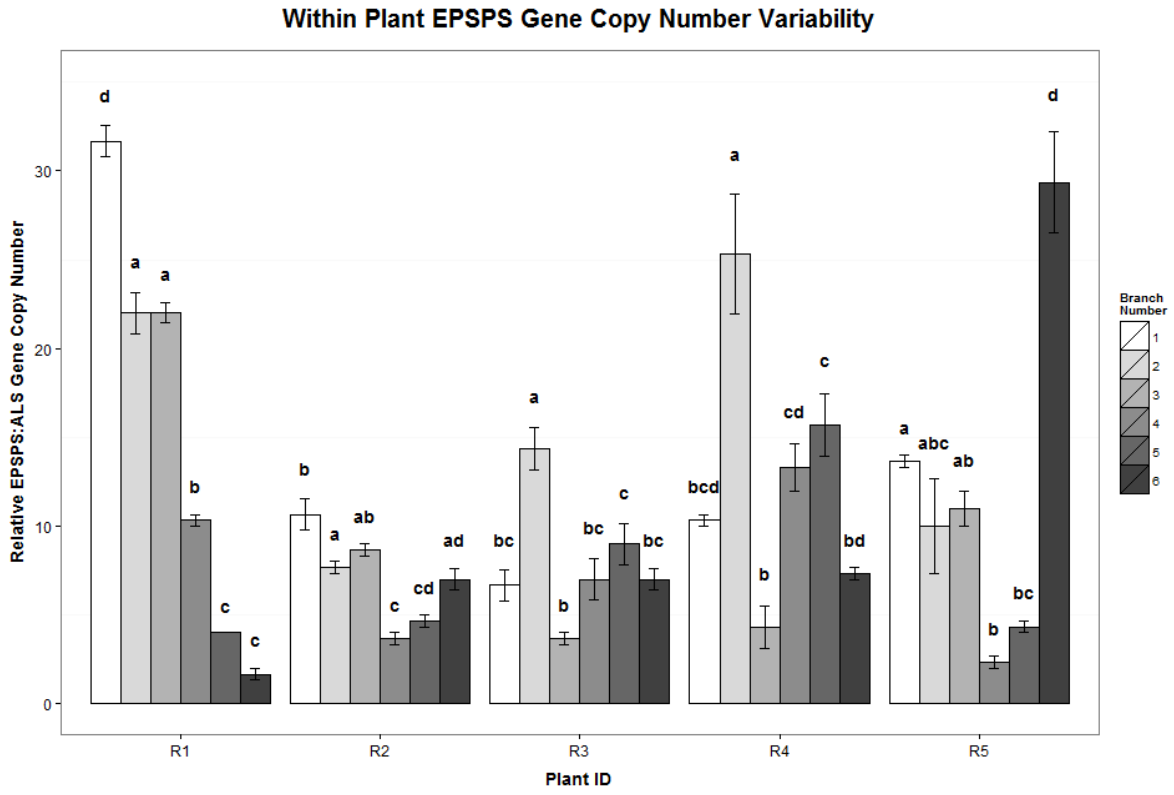


Figure 3.4. *EPSPS* copy number differences between the side shoots of five separate Palmer amaranth glyphosate resistant plants (R1-R5). Mean ± SE for each clone with Tukey HSD test results shown as letters above columns. Means not sharing the same letter are significantly different. ANOVAs/Tukey HSDs were run on each plant separately to analyze within plant significant differences in *EPSPS* gene copy number.

REFERENCES

Beckie H (2011) Herbicide-resistant weed management: focus on glyphosate. Pest Manag Sci 67: 1037-1048

Chandi A, Milla-Lewis SR, Giacomini D, Westra P, Preston C, Jordan DL, York AC, J.Burton JD, Whitaker JR (2012) Inheritance of evolved glyphosate resistance in a North Carolina Palmer amaranth (*Amaranthus palmeri*) biotype. Int J Agronomy doi:10.1155/2012/176108

Cristoffers MJ, Varanasi AV (2010) Glyphosate resistance: genetic basis in weeds. Pages 119-140 In Nandula VK (ed) Glyphosate resistance in crops and weeds: History, development and management. Hoboken NJ: Wiley. 344 p

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183: 557-564.

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang JM, Preston C, Leach JE, Westra P (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proc Natl Acad Sci USA 107:1029–1034

Gaines TA, Shaner DL, Ward SM, Leach JE, Preston C, Westra P (2011) Mechanism of resistance of evolved glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*). J Ag Food Chem 59:5886-5889

Ge X, d'Avignon DA, Ackerman JJ, Sammons RD (2010) Rapid vacuolar sequestration: the horseweed glyphosate resistance mechanism. Pest Manag Sci 66:345-348

Giacomini D, Westra P and Ward SM (2014) Impact of genetic background in fitness cost studies: an example from Palmer amaranth. Weed Sci 62:29-37.

Heap I (2013) The International Survey of Herbicide-Resistant Weeds. http://www.weedscience.org. Accessed October 21, 2013

Jasieniuk M, Brule-Babel AL, Morrison IN (1996) The evolution and genetics of herbicide resistance in weeds. Weed Sci 44:176-193

Jasienuik M, Ahmad R, Sherwood AM, Firestone JL, Perez-Jones A, Lanini WT, Mallory-Smith C, Stednick Z (2008) Glyphosate-resistant Italian ryegrass (*Lolium multiflorum*) in California: distribution, response to glyphosate, and molecular evidence for an altered target enzyme. Weed Sci 56:496-502

Kniss AR, Streibig JC (2015) Statistical Analysis of Agricultural Experiments with R. Linear Regression. http://rstats4ag.org/. Accessed: February 2, 2015

Mohseni-Moghadam M, Schroeder J, Ashigh J (2013) Mechanism of resistance and inheritance in glyphosate resistant Palmer amaranth (*Amaranthus palmeri*) populations from New Mexico, USA. Weed Sci 61:517-525

Ng CH, Ratnam W, Surif S, Ismail BS (2004) Inheritance of glyphosate resistance in goosegrass (*Eleusine indica*). Weed Sci 52:564-570

Neve P (2007) Challenges for herbicide resistance evolution and management: 50 years after Harper. Weed Res 47:365-369

Powles SB (2010) Gene amplification delivers glyphosate-resistant weed evolution. Proc Natl Acad Sci USA 107:955–956

Pratley J, Urwin N, Stanton R, Baines P, Broster J, Cullis K, Schafer D, Bohn J, Krueger R (1999) Resistance to glyphosate in *Lolium rigidum*. I. Bioevaluation. Weed Sci 47:405–411

Preston C, Wakelin AM (2008) Resistance to glyphosate from altered herbicide translocation patterns. Pest Manag Sci 64:372–376

Ribeiro DN, Dayan FE, Pan Z, Duke SO, Shaw DR, Nandula VK, Baldwin BS (2011) EPSPS gene amplification inheritance in glyphosate resistant *Amaranthus palmeri* from Mississippi. Page 137 *in* 2011 Proceedings of the Southern Weed Science Society. Las Cruces, NM: Southern Weed Science Society

Shaner D, Nadler-Hassar T, Henry WB, Kroger C (2005) A rapid in vivo shikimate accumulation assay with excised leaf discs. Weed Sci 53:769–774

Simarmata M, Bughrara S, Penner D (2005) Inheritance of glyphosate resistance in rigid ryegrass (*Lolium rigidum*) from California. Weed Sci 53: 615–619.

Teaster ND, Hoagland RE (2014) Genomic stability of Palmer amaranth plants derived by macro-vegetative propagation. Am J Plant Sci, 5: 3302.

Vila-Aiub MM, Goh SS, Gaines TA, Han H, Busi R, Yu Q, Powles SB (2014) No fitness cost of glyphosate resistance endowed by massive EPSPS gene amplification in Amaranthus palmeri. Planta 239:793-801

Wakelin AM, Preston C. (2006) Inheritance of glyphosate resistance in several populations of rigid ryegrass (*Lolium rigidum*) from Australia. Weed Sci 54:212–219.

Wicker T, Yahiaoui N, Keller B. (2007) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J 51: 631–641.

Wiersma AT, Gaines TA, Preston C, Hamilton P, Giacomini D, Robin Buell C, Leach JE, Westra P. (2015) Gene amplification of 5-enol-pyruvylshikimate-3-phosphate synthase in glyphosate-resistant Kochia scoparia. Planta 241:463-474.

Yu Q, Cairns A, Powles S (2007) Glyphosate, paraquat and ACCase multiple herbicide resistance evolved in *Lolium rigidum* biotype. Planta 225:499-513.

Yuan Z, Zhang L, Zhang L, Chen W, Yan Z, Zheng Y, Zhang H, Yen Y. (2011) Mitotic illegitimate recombination is a mechanism for novel changes in high-molecular-weight glutenin subunits in wheat-rye hybrids. Plos One 6(8): e23511.

CHAPTER 4. SEQUENCE ASSEMBLY AND CHROMOSOMAL LOCALIZATION OF THE *EPSPS* AMPLICON IN PALMER AMARANTH

4.1 INTRODUCTION

The discovery of a gene amplification-based mechanism of herbicide resistance came in 2010 with the confirmation of 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) gene duplication in *Amaranthus palmeri*, more commonly known as Palmer amaranth (Gaines et al., 2010). Although this was the first documented case of herbicide resistance via naturally-occurring gene copy number increase in weeds, it was not a surprising finding, as this mechanism of resistance had already been well documented in insecticide resistance (Devonshire and Field, 1991) and glyphosate-treated plant cell tissue cultures (Shyr et al., 1992). However, the immense scale of the duplications was unusual, with many resistant individuals exhibiting 160+ copies of the *EPSPS* gene. Understanding the mechanism behind this massive amplification will not only help to inform weed scientists of the genetic background and selection pressures that produce this type of herbicide resistance but also advance the understanding of gene amplification mechanisms in the basic plant science community.

So far, this form of glyphosate resistance has been found in four other weed species, including *Amaranthus tuberculatus*, *Lolium multiflorum*, *Kochia scoparia*, and *Amaranthus spinosus* (Sammons and Gaines, 2014). None of these species exhibit as many *EPSPS* copies as seen in Palmer amaranth, but *A. spinosus* has up to 37 copies and is known to intercross with Palmer (Gaines et al., 2012) and likely gained the resistance trait via interspecific hybridization (Nandula et al., 2014). Exactly why Palmer amaranth can acquire so many

copies is still unknown, but it has been suggested that Palmer is comparatively further along in its evolution of resistance and the other species will accumulate similarly high numbers of *EPSPS* copies over time (Wiersma et al., 2015). In any case, this level of single gene duplication is rare in plants and the gene duplication mechanism is still unknown. Elucidating this mechanism is a primary goal of this work with the hopes that it will reveal more about the origin and evolution of glyphosate resistance in Palmer amaranth.

Towards this goal, we have sequenced the *EPSPS* gene from a Palmer amaranth plant originating from one of first identified glyphosate resistant populations out of Macon County, Georgia. The main objective of this project was to sequence and assemble the *EPSPS* amplicon using a multi-sequencing platform approach in which the shortcomings of each sequencing technology were counterbalanced by one another. This assembled amplicon was searched for any features that would indicate how gene duplication occurred and a fluorescent in situ hybridization (FISH) assay was used to confirm the likelihood of unequal recombination as the mechanism of *EPSPS* copy number gain and loss.

4.2. MATERIALS AND METHODS

Fosmid Assembly.

Genomic DNA was extracted from a glyphosate resistant Palmer amaranth plant with multiple *EPSPS* copies (*EPSPS* copy number = 122, determined via qPCR as described in Gaines et al. 2014). The gDNA was randomly sheared by repeated pipetting and was run on a 0.7% agarose gel (1V/cm, 16 hours) to size select for 20-45kb bands. The size selected gDNA was gel extracted using a QIAquick Gel Extraction kit (QIAGEN) then ligated into

fosmid vectors (pSMART FOS, CopyRight v2.0 Fosmid Cloning Kit, Lucigen), packaged into phage particles (MaxPlax Lambda Packaging Extracts, Epicentre), and transfected into Replicator FOS Cells, following the kit protocol. Cells were plated onto a 20 x 20cm Bio-Assay plate and lifted onto a cellulose membrane for *EPSPS* probe hybridization to identify *EPSPS* positive clones. These clones were confirmed to contain *EPSPS* via PCR of the first and last exon (exon1_F: 5'- CCAAAACCCAGTTACCCAAA-3'; exon1_R: 5'-GACCCAGGCAATTGAACAGT-3'; exon8_F: 5'- GGCAACAGTTGAGGAAGGAT-3'; exon8_R: 5'-TCCCGGATCAAGGATAGTGA-3'), and then sent off for sequencing on the Illumina MiSeq and PacBio platforms. More than 400 fosmids were sequenced and assembled with the assembly program PCAP (Huang et al. 2003). Prior to assembly, all fosmid vector sequences were removed from the reads using the Cutadapt program (available at https://code.google.com/p/cutadapt/). The five longest fosmids were manually finished using Consed (Gordon et al. 1998) and aligned in CLC Genomics.

Amplicon Extension.

Three additional genomic sequence data sets from *Amaranthus palmeri* were generated and used for extension out from the fosmid consensus sequence. Both paired-end and 20kb mate-pair Illumina libraries were prepared and sequenced at the Monsanto Genome Analysis Center using Nextera Library Preparation Kits (Illumina) and a MiSeq sequencing platform (Illumina). The mean read length (after adapter and quality trimming via cutadapt) was 92nt for both the paired-end and mate-pair data sets and overall genome coverage was 67x and 55x for the paired-end and mate-pair data sets, respectively.

Coverage was calculated using the Lander/Waterman equation (Lander and Waterman, 1988):

$$C = LN / G$$

in which C is the coverage, L is the mean read length, N is the number of reads, and G is the haploid genome length (533 Mb for Palmer amaranth). The third data set was also generated with the Illumina platform, but used the TruSeq Synthetic Long-Read DNA Library Prep Kit for library preparation. Long read assembly was performed by Illumina, resulting in an average read length of 2,692nt (maximum read length of 24,313nt) and 15x coverage.

The CLC mapper module from CLC Assembly Cell v.4.0 was used for reference mapping of all data sets. Illumina paired end reads were mapped to the longest fosmid build with mapping parameters set to 90% similarity and 80% length fraction (clc_mapper -s 0.9 –l 0.8 –q –p fb ss 100 500). Mismatch, insertion, and deletion costs were kept at the default settings of -2, -3, and -3, respectively, and non-specific reads were mapped at random. Similar mapping parameters were also used for the TruSeq long reads and 20kb mate pair reads data sets, expect for the paired read parameters; TruSeq reads were mapped singly while the 20kb mate paired reads were mapped as mate pairs using gap parameters of 15000-25000nt (clc_mapper –s 0.9 –l 0.8 –q –p bf ss 15000 25000).

SNP Detection.

Single nucleotide polymorphism (SNP) calling was accomplished using the CLC Genomics platform tool SNP Detection, which operates under a Neighborhood Quality Standard algorithm based on Altshuler et al. (2000). The SNP calling parameters included a

window length of 11, within which the maximum number of gaps and mismatches was restricted to 2, the minimum quality of the central base was 20, and the minimum average quality of the surrounding bases was 15. For a SNP to be called, the minimum coverage had to be at least 10 reads deep for that site. The Palmer amaranth plant chosen for this sequencing project had 122 *EPSPS* gene copies, as determined by qPCR, so to detect differences in a single one of these copies, the minimum variant frequency was lowered to 0.8% (1/122 = 0.00819).

Repeat Search and ORF Annotation.

Repeats were found with the UGENE v1.11.4 program from Unipro (Okonechnikov et al, 2012). Using the Find Repeats tool in this software package, both direct and inverted repeats were identified ranging in length from 50bp up to 1000bp. Only repeats with 100% identity were extracted in the initial analysis. These repeat sets were grouped by length (50bp, 100bp, 500bp, and 1000bp) and orientation (direct and inverted), then exported as a comma-separated value file. This data was fed into a Processing script modified from Wattenberg (2002) by Szczesny (2008) to create an arc diagram to visually represent the repeat structures in this build. A dotplot was also created using the PipMaker program (Schwartz et al., 2000), comparing the build against itself to call any repeated regions.

Open reading frames (ORFs) were found using the ORF Finder online utility from NCBI. This program searched for ORFs with a minimum length of 300bp (100 amino acids) and the resulting translated proteins were analyzed on the NCBI BLAST webserver

(blastp). Only hits with an alignment score greater than 50 and e-value less than 0.1 were considered significant and only the top significant hit for each ORF is reported here.

Allele-Specific PCR and Non-Duplicated EPSPS Assembly.

Allele-specific PCR.

Previous SNP analysis (Giacomini et al., 2015) revealed four closely grouped low-frequency SNPS co-occurring on the same reads in exon 5 of the *EPSPS* gene. Four primers were created across this SNP-dense region, two of which aligned to the majority of the reads which matched the duplicated *EPSPS* gene (Exon5noSNP_F1: 5'-TGAGGGTGATGCTTCAAGTG-3'; Exon5noSNP_R1: 5'-CACTTGAAGCATCACCCTCA-3'). The other two primers aligned to the reads containing the low frequency SNPs which putatively matched the non-duplicated *EPSPS* gene (Exon5SNP_F1: 5'-AGAGGGGGACGCTTCTAGTG-3'; Exon5SNP_R1: 5'- CACTAGAAGCGTCCCCCTCT -3'). Two more primers were created at both ends of the *EPSPS* gene primer in regions of the 5'UTR and 3'UTR where no SNPs occurred (5'UTR_F1: 5'-CCACTTTCTCTTTGCCCACC-3'; 3'UTR_R1: 5'-CAAAACCTTCGGCGTACAAT-3'). Allele-specific PCR was run on cDNA extracted from a glyphosate resistant Palmer amaranth plant using EmeraldAmp polymerase and the following thermoprofile: one minute at 98°C, 35 cycles of 98°C for 15 seconds, 60°C for 30 seconds, and 72°C for 8 minutes, followed by a final extension at 72°C for 8 minutes. Bands were extracted from the gel and purified using the Qiagen QIAquick Gel Extraction Kit. The PCR products were cloned into pCR®-XL-TOPO® plasmid vectors and transformed into OneShot® TOP10 chemically competent cells using the TOPO® XL PCR Cloning kit (Life Technologies, K475020). Plasmids were checked for the correctly sized insert through restriction

enzyme analysis and sent off for sequencing at the CSU Proteomics and Metabolomics Facility. The resulting sequences were assembled, confirming the previously called SNPs and generating two different *EPSPS* alleles.

Non-duplicated EPSPS Fosmid Assembly and Analysis.

From the already generated fosmids described above, non-duplicated *EPSPS*-containing fosmids were selected using the allele-specific PCR primers. These fosmids were sequenced and assembled as before, using the PCAP and Consed programs to generate a 54,139bp-long consensus sequence. This sequence was aligned the duplicated *EPSPS* allele consensus in UGENE using the ClustalW multiple sequence alignment algorithm and sequence similarity calculations were performed on this alignment using UGENE's distance matrix tool (distance algorithm: identity; profile mode: percents). Repeat searches and ORF annotations were also run on this non-duplicated *EPSPS* allele sequence to determine the composition of the flanking regions.

Fluorescent In Situ Hybridization.

Chromosome spreads.

The following protocol is an adaptation of the methods found in Kato (1999). Starting with Georgia glyphosate resistant and susceptible Palmer amaranth that was grown in the greenhouse under 400-W sodium halide lamps (14-hour photoperiod and temperatures of 24 C day/18 C night), side shoots were cut off of the main stem of a 12-inch plant and dipped into rooting hormone. The shoots were then planted into moist soil, covered with a plastic bag to maintain a humid environment, and grown for 7 days under

continuous light at room temperature (25°C). The plants were then carefully extracted from the soil and washed by placing the roots into a beaker with water and gently swirling until the dirt came loose. The ends of the roots (1-2 cm) were cut with a razor, gently transferred to a 0.65mL tube with a hole poked in the lid, and immediately placed into a pressurized nitrous oxide (N_2O) chamber at 10.5 atmospheres for 2.5 hours. The N_2O treatment was done to arrest cells in the metaphase stage of mitosis because N_2O prevents spindle fiber attachment, blocking the progression of the cell cycle to anaphase.

After venting the N_2O chamber and removing the tubes, 90% cold acetic acid was immediately added to each tube to fix the chromosomes and prevent resumption of the cell cycle. Tubes were held at 4°C for 30 minutes and then the acetic acid was exchanged for cold 70% ethanol. The ethanol was then exchanged for cold 1x citrate buffer (50 mM sodium citrate, 50 mM EDTA, pH adjusted to 5.5 using citric acid) and the tubes were held at 4°C for 60 minutes. Once the roots were equilibrated with the citrate buffer, they were removed and placed onto filter paper. Using a scalpel, the root meristem was carefully cut off and placed into a new 0.5mL tube along with 50 µL root tip digestion mix (1% w/w Pectolyase Y-23 and 2% w/w Cellulase Onozuka R-10 in 1x citrate buffer). Root tips were digested at 37°C for 65 minutes, and then transferred to ice to halt the digestion.

The digestion mix was replaced with 70% ethanol (exchanging out the 70% ethanol two more times to eliminate all of the enzyme mix), and then replaced with 30 µL of 3:1 acetic acid:methanol. A pipette tip was used to mash the root tips against the side of the tube and then 8 µL of the digested cells was dropped onto the center of a clean slide. A few slides were stained with DAPI and viewed at 100x to verify good spreads and do initial

chromosome counts. The rest of the slides were cross-linked at 120 mJ and used immediately for hybridization (see *FISH Probe Hybridization*).

EPSPS FISH Probe Construction.

The entire *EPSPS* gene including introns was used as a probe for this FISH assay. Primers designed against the first and last codon (EPS_F1: 5'- GGCTCAAGCTACTACCATC AACAATGG-3' and EPS_R1: 5'- TCAAGGATAGTGACGGGAACATCTGC-3') were used to PCR amplify the 10,160bp product. The band was excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen). This product was then nick-translated and fluorescently labeled using the following protocol: Mixed 10 µL DNA (200 ng/µL purified *EPSPS* PCR product) with 2 µL NEB buffer 2 (New England Biolabs), 2 µL non-labeled dNTPs (2 mM each, mixed), and 0.5 µL fluorescently labeled dNTP (Texas Red-5-ddATP, Perkin Elmer). To this, 10 µL of DNA polymerase I (10 U/µL) and 0.4 µL DNase I (100 mU/µL) was added, mixed by pipetting, and placed at 15°C for two hours. To stop the reaction, 2 µL of 0.5M EDTA was added and the probe was stored in the dark at -20°C until use. This FISH probe labeling method and the following FISH assay protocol are based on Kato et al. (2006).

FISH Probe Hybridization.

Five µL of salmon sperm DNA was pipetted onto the center of each chromosome spread slide and a plastic cover slip was laid over the top. The slides and the probe DNA (1.0 µL Texas red-labeled probe in 4.0 µL 2x SSC TE for each FISH slide) were denatured in a boiling water bath for 5 min (100°C). The slides and probe were quickly transferred to a

metal tray on ice to quickly cool for two minutes, then 5 µL of the probe was dropped onto the center of each slide cell suspension. Slides were placed in a humid storage container overnight at 55°C, and then moved into a Coplin jar containing 2x SSC to remove the cover slip. Slides were then transferred to another Coplin jar containing 55°C 2x SSC and stored for 20 minutes at 55°C. Once this wash step was done, the slides were removed and one drop of Vectashield mounting medium with DAPI was applied to each cell spread and a cover slip was placed over the top. Slides were viewed under 100x oil immersion on a Zeiss Axiovert 200M inverted fluorescence microscope (Carl Zeiss, Gottingen, Germany) containing a DAPI filter set (D360/40-nm excitation wavelength, 400-nm dichroic long pass mirror, D460/50-nm emission wavelength) and a FITC filter set (HQ480/40-nm excitation wavelength, 505-nm long pass dichroic mirror, HQ535/50-nm emission wavelength).

4.3. RESULTS AND DISCUSSION

The EPSPS Amplicon is >110kb in Length.

More than 400 fosmids containing the *EPSPS* gene were sequenced and assembled with the PCAP assembly program (Huang et al. 2003), resulting in 5722 contigs with an N50 of 12213. Forty-six of the assembled fosmids were over 30kb in length and six of these long fosmids were manually finished using Consed (Gordon et al. 1998). The fosmids chosen for manual finishing were selected to cover as far out from the *EPSPS* gene as possible, resulting in 19.4kb of additional sequence assembled upstream and 10.5kb assembled downstream. Sequence similarity calculations between the fosmids indicated 100% identity between *EPSPS* gene copies and the surrounding sequence (Figure 4.1).

Mapping of the Illumina paired-end reads to this 40kb build confirmed the fosmid consensus and again uncovered very little sequence diversity across the *EPSPS* gene copies. Single nucleotide polymorphism (SNP) analysis on the paired-end mapped reads called 380 SNPs across the entire build, totaling less than 1% of the sequence. Two sites of sequence divergence were found in this mapping, appearing on both sides of the gene, approximately 3kb down and 8.6kb upstream from *EPSPS* (Figure 4.2 A). At the upstream divergence site, 60% of the reads completely matched the fosmid consensus while the other 40% matched up till 10,893bp from the beginning of the fosmid build, then deviated into a unique new downstream consensus (Figure 4.2 B). The downstream divergence site showed the opposite pattern, with 58% of the reads matching the fosmid consensus and the rest diverging into a new upstream consensus (Figure 4.2 B).

Mapping of the paired-end reads also revealed no drop in read coverage across the amplicon consensus sequence (Figure 4.3 A). If the ends of the amplicon had been reached, the flanking sequence would no longer be identical (extending out into unique genomic regions) and the read coverage across the ends of the consensus sequence would drop sharply. As this did not happen, it was determined the amplicon ends had not been found. Instead, the coverage rose across many regions of the mapped build, indicating the presence of repeated elements. At both divergence sites, read coverage was doubled and remained high out to the ends of the build. At the very ends of the consensus, coverage increased to more than 6x that of the *EPSPS* gene. When repeat searches were run on the build, a 2kb-long repeat was found at both ends (red arrows, Figure 4.2) and the very high coverage across those repeats suggested they occurred many times either within the *EPSPS* amplicon or elsewhere in the Palmer amaranth genome. BLAST searches of this repeat

element revealed no clear identity and no evidence that these repeats were some type of transposable element. PacBio sequencing of the fosmids made assembly of these flanking 2kb repeats possible, resulting in seven 2kb repeats downstream and at least two 2kb repeats upstream. Unfortunately, the selected fosmids did not capture sequence beyond these repeats.

To get past the flanking repeats and further extend the build, whole genome Illumina TruSeq Synthetic Long-Reads and Illumina 20kb mate pair reads were mapped to the fosmid consensus. As in the short read data set, the TruSeq reads showed the same two divergence sites as well as increased read coverage beginning out from these two sites. This increased coverage indicated that this portion of the build was either present elsewhere in the genome or was duplicated within the amplicon. The latter hypothesis was confirmed by the 20kb mate pairs which placed the duplicated unit both before and after the *EPSPS* gene (Figure 4.3 B). Additional extensions to the build were possible using the much longer TruSeq reads, accomplished by pooling all reads that extended off the ends of the amplicon build, generating a consensus from these reads to be concatenated onto the ends of the previous build, and then remapping all of the reads again to the new build. Through this iterative mapping process, an additional 31.2kb of sequence was added on, generating an amplicon build of 110,307bp. Though this entire consensus was supported by the long TruSeq reads as well as PCR/sequencing across this region (data not shown), the last 25kb of the build was not confirmed by the 20kb mate pairs (Figure 4.3 B). The Illumina paired end read coverage was also much higher across this region (Figure 4.3 A).

Repeat Elements Make up the Majority of the Amplicon.

The repeat discovery utility in UGENE was used to find all repeated sequences within the 110kb amplicon. These repeats were visualized using both a Processing 2 script to draw arcs connecting repetitive regions as well as a dotblot highlighting any portion of the sequence that had 100nt of sequence that was 100% identical to another region within the build (Figure 4.4). This repeat analysis revealed the majority of the sequence surrounding the *EPSPS* gene was filled with repetitive elements, both direct and inverted. The long direct repeats include the 28.3kb repeat detailed above, as well as the 2kb-long tandemly arranged repeats within that longer repeat, present in tandem sets of seven. At least three more of these 2kb repeats are evident at the very beginning of this build, but the repetitive nature of this sequence prevented further upstream assembly. Two more long directly repeated regions (1088bp and 2595bp in length) were also observed downstream, encompassing the last 25kb of the build. The inverted repeats present in this amplicon build are much shorter than the direct repeats. Two of the largest inverted repeats occur on either side the *EPSPS* gene and bear some resemblance to MITEs (miniature inverted-repeat transposable elements), including the 8-10bp target site duplications (TSDs) that are diagnostic for MITEs and other types of DNA transposons (Figure 4.5).

Alongside the repeat analysis, open reading frame (ORF) discovery using UGENE was also run, finding very few ORFs within the build (Figure 4.6 A). A previously described transposon (Gaines et al. 2013) was present twice in the sequence on both sides of the *EPSPS* gene. Another group of ORFs was present towards the end of the build, encoding for two peptides with significant matches to the Pfam family of plant mobile domains (PMD, PF10536) and one peptide with homology to a GRF zinc finger DNA binding domain (zf-

GRF, PF06839). Two short ORFs were discovered within the 2kb repeat elements, but BLAST searches against the NCBI nr database and Pfam protein family database returned zero significant hits.

Although the ORF search returned a few elements with similarity to mobile elements, it is still not proven that these were involved in the duplication mechanism. All of these occurred within the amplicon at the same coverage as the rest of the sequence, so it is possible these elements inserted into the flanking regions and were taken along for the ride when the first duplication occurred. More interesting is the presence of the highly structured repeats surrounding of *EPSPS* gene. The long tandem repeats have been noted in other examples of gene duplication, including plant disease resistance genes (Leister, 2004), human genomic disorders (Shaw and Lupski, 2004), and rDNA repeats (Eickbush and Eickbush, 2007). These examples acknowledge unequal recombination as the driving force of gene duplication and loss. As demonstrated in Giacomini et al. (2015), *EPSPS* gene number can change drastically not only between reproductive generations of Palmer amaranth propagation, but also within an individual plant. This level of gene copy number gain and loss is unlikely to be attributable solely to the action of transposable elements for a number of reasons:

(1) Transposons clip and copy genes one at a time, so an unprecedented level of transposon activity would be required to occur during mitosis and meiosis to reach this amount of gene duplication and loss.

(2) The presence of introns and flanking regions in the duplicated sequence indicates a DNA transposase is responsible for the duplicated genes. These types of transposons operate under a “cut-and-paste” duplication mechanism wherein any increase

in gene copy number must occur within the replication fork during the S phase of the cell cycle. This short window for gene duplication to occur further limits the likelihood that a transposon is guiding this process.

(3) With this work the current length of the duplicated *EPSPS* amplicon is greater than 110kb. This is nearly a log scale difference in sequence length from other known transposon-mediated sequence duplications (Feschotte and Pritham, 2007).

This is not to say that transposons are not involved at all in this mechanism. It has been hypothesized that the action of transposons can promote chromosomal duplications, deletions, inversions, and translocations via either recombination between repeated transposable elements or alternative transposition (Gray, 2000). It could be that one or a few transposable element-mediated events resulted in the *EPSPS* gene becoming inserted into a novel genomic position that was repeat rich and susceptible to slippage during mitosis and meiosis, leading to tandem duplications of the gene and expansion of the tandem repeats with each new cycle of cell division. However, the massive gene gain and loss is likely due to unequal recombination. Confirming the orientation and order of the *EPSPS* repeats on the chromosomes will help test this hypothesis.

A Non-Duplicated EPSPS Allele.

The SNP analysis revealed a number of low frequency SNPs that all grouped onto the same set of reads in the Illumina paired-end data set. These appeared to represent a second *EPSPS* allele, so allele-specific PCR and direct sequencing was carried out to confirm this. The second allele differed from the duplicated *EPSPS* allele by 19 exonic SNPs, three of which encoded an amino acid change (Figure 4.7). Two of the amino acid changes occurred

in the chloroplast transit peptide portion of the sequence while the other non-synonymous SNP occurred in exon 4, encoding a lysine to arginine change (K289R). The second allele was confirmed to be occurring only once in the genome (non-duplicated) via quantitative PCR of gDNA from multiple resistant plants (data not shown).

Fosmids containing the second allele were generated, sequenced, and assembled to produce a 54kb-long contig. Comparisons between the duplicated *EPSPS* allele and the non-duplicated second allele confirmed the previously discovered SNPs along with an additional 1859 SNP and indel differences in the intronic and untranslated regions (Figure 4.8). Total percent identity between the two allele sequences was 92%, excluding gaps. The Beyond the UTRs, sequence similarity between the two alleles dropped off sharply, indicating very different flanking regions. Unlike the repeat-heavy flanking sequence surrounding the duplicated *EPSPS* allele, the non-duplicated allele was surrounded by a number of long ORFs. BLAST analyses of these ORFs identified most of them as retrotransposons with high sequence similarity to the Ty/Gypsy family of retrotransposable elements. Two ORFs close to the *EPSPS* gene were identified as part of a FAR1 (FAR-RED IMPAIRED RESPONSE 1) gene and another long ORF downstream of that gave high scoring hits to the UDP-D-glucuronate 4-epimerase 3 gene (Figure 4.6 B).

This sharp divergence in sequence between the two alleles suggests they are located at two separate loci in the Palmer amaranth genome. This data indicates the duplicated *EPSPS* gene was moved to a different genomic location. Interestingly, the sites where sequence similarity between the two alleles drop off also mark the two spots where the two long inverted repeats resembling MITE elements are found in the duplicated *EPSPS* sequence. It may be that these putative MITEs were involved in the movement of the

EPSPS gene to the repetitive genomic region detailed above. Once the gene was nested in this repeat-rich sequence, other mechanisms of gene duplication (and loss) acted upon the gene, leading to the extremely high *EPSPS* copy number genotypes seen in many glyphosate resistant Palmer amaranth plants today.

The EPSPS Duplicated Genes are arranged in Tandem on a Single Chromosome.

The fluorescent in situ hybridization (FISH) assay revealed a single chromosomal location of the *EPSPS* genes in a high copy number glyphosate resistant individual (Figure 4.8). Although there is not yet sequence-based confirmation of a tandem arrangement of *EPSPS* gene copies, this result would indicate the duplicate genes are very near one another and likely in tandem. This result contradicts earlier FISH assays (Gaines et al., 2010), so more FISH work is needed to confirm this result.

Conclusion.

The presence of the duplicated EPSPS gene in the middle of a highly repetitive portion of sequence completely absent of any other genes (besides transposons) is a striking finding, especially since this data shows the entire 110+kb of this build is duplicated along with the *EPSPS* gene. Duplication on this scale results in an extraordinary increase in genome size with each gain and loss of *EPSPS* copies. For the Palmer amaranth plants that have 160 *EPSPS* copies, this would mean these duplications give rise to at least an extra 17.6Mb of sequence or more than 1.6% of the entire Palmer amaranth genome (533 Mb/1C). The plants appear to bear this substantial gain to their genome well, with no fitness cost associated with increased *EPSPS* copy number (Giacomini et al., 2014; Vila-Aiub

et al., 2014). Adding to this puzzle, it is now known that Palmer amaranth is also able to quickly change its *EPSPS* gene copy number both across generations and within a single plant. These results taken together seem to indicate a novel duplication mechanism is at work and although it is still not known exactly what this mechanism is, the findings from this project should help guide future research efforts.

4.4 FIGURES

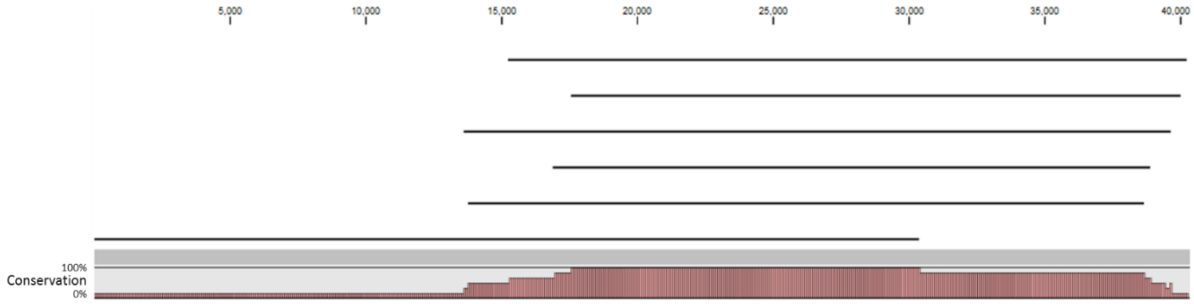


Figure 4.1. Alignment of finished *EPSPS* fosmids. Conservation scores at each nucleotide site are shown as pink bars below the alignment. The fosmid consensus assembly length is 40,217nt.

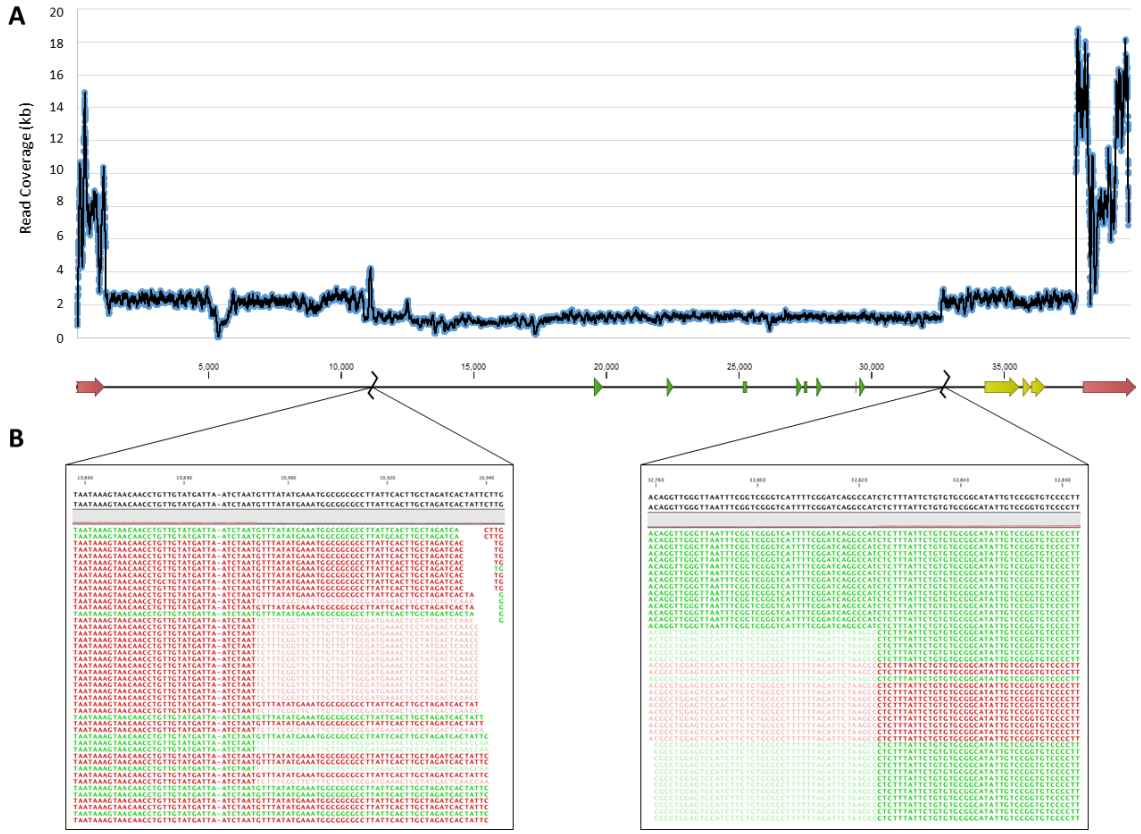


Figure 4.2. Assessment of read coverage and sequence divergence across the *EPSPS* fosmid consensus. A. Paired end Illumina 100bp read coverage across the *EPSPS* fosmid consensus. Annotations for the *EPSPS* gene (green arrows indicating the eight *EPSPS* exons), Activator transposase (yellow arrows), and flanking repeats (red arrows) are indicated on the sequence at the bottom. Divergence sites that flank the *EPSPS* gene are indicated on the sequence and shown in detail in B.

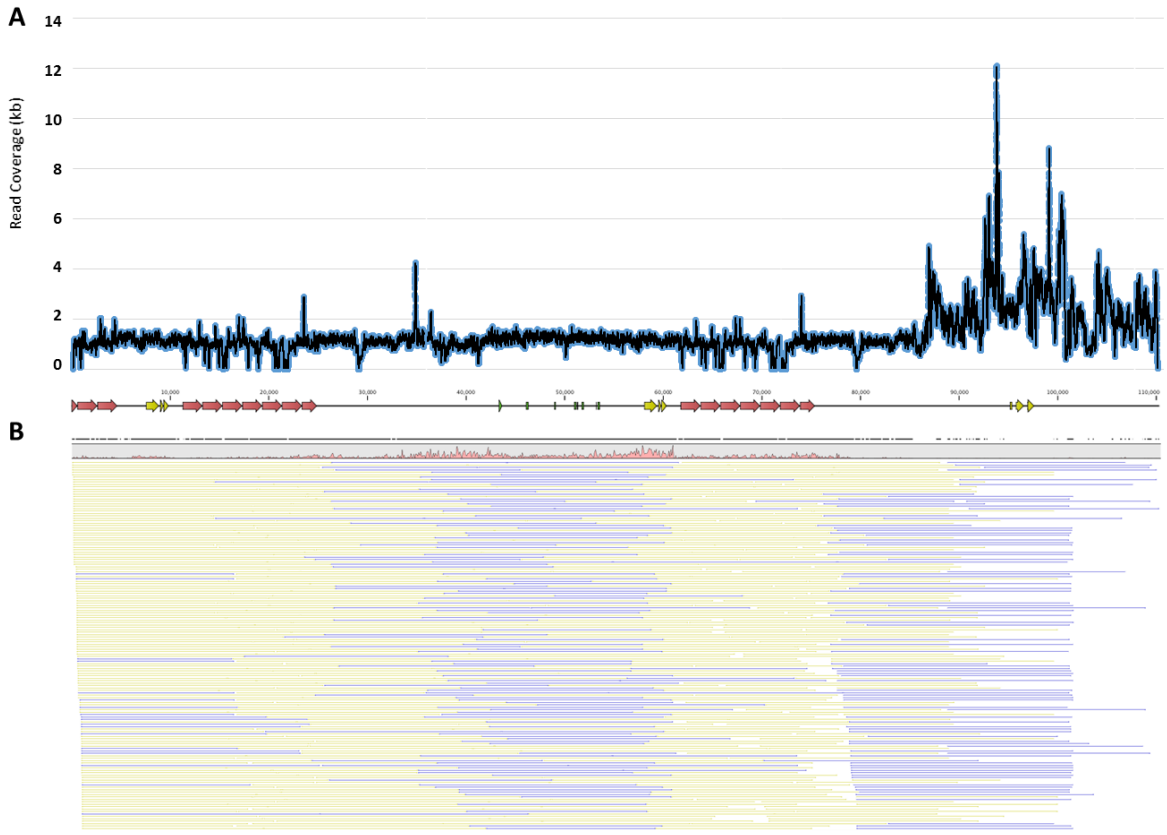


Figure 4.3. Read coverage across the extended *EPSPS* build. A. Paired end Illumina 100bp read coverage across the extended build. Annotations for the *EPSPS* gene (green arrows indicating the eight *EPSPS* exons), open reading frames (yellow arrows), and flanking repeats (red arrows) are indicated on the sequence at the bottom. B. Mapping of the 20kb Illumina mate paired reads to the extended build. Blue lines indicate uniquely mapped mate pairs and yellow lines indicate mate pairs that may map to multiple locations in the build. Each line indicates the distance between the two 100bp mate-paired reads (15-25kb).

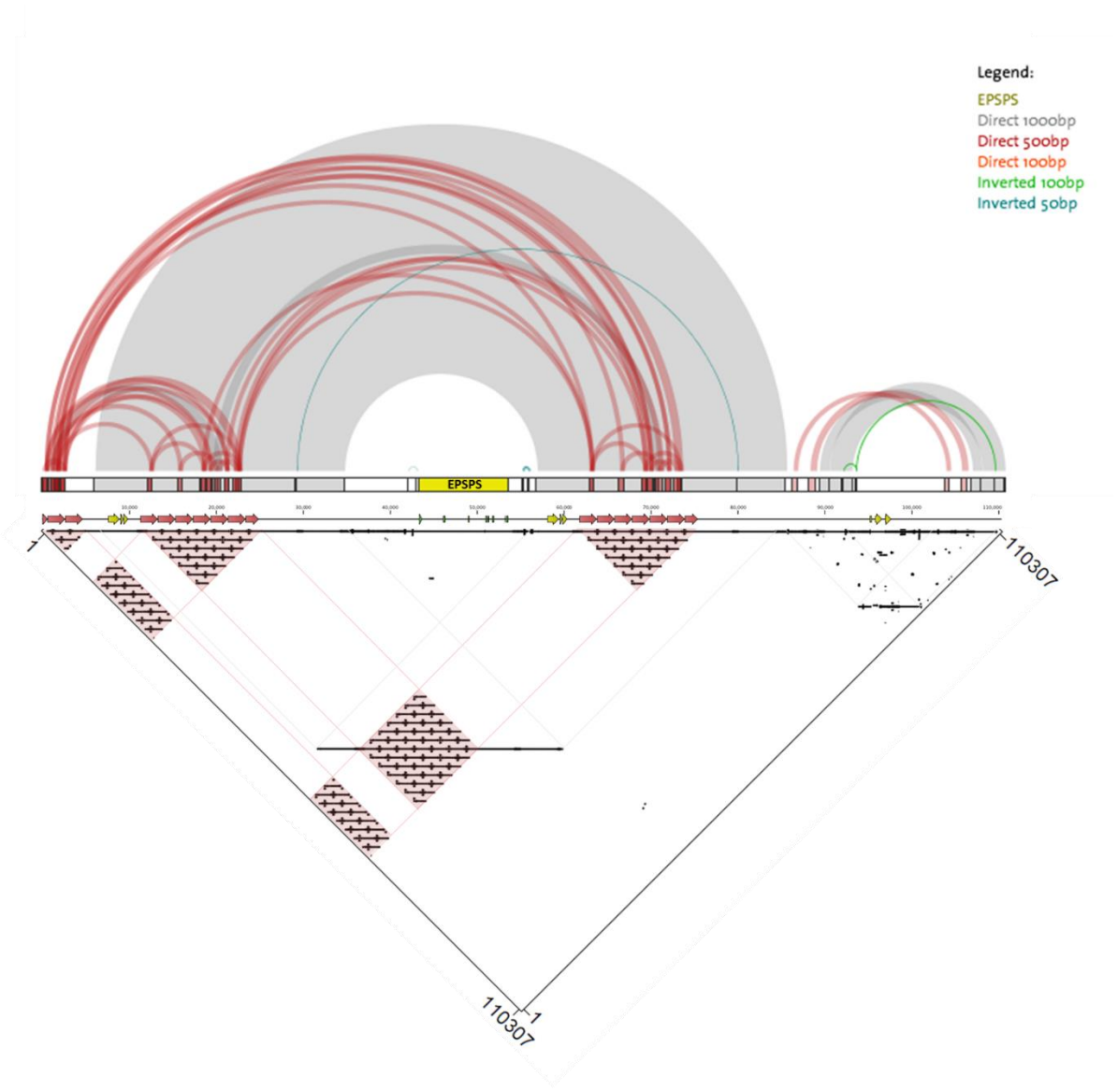
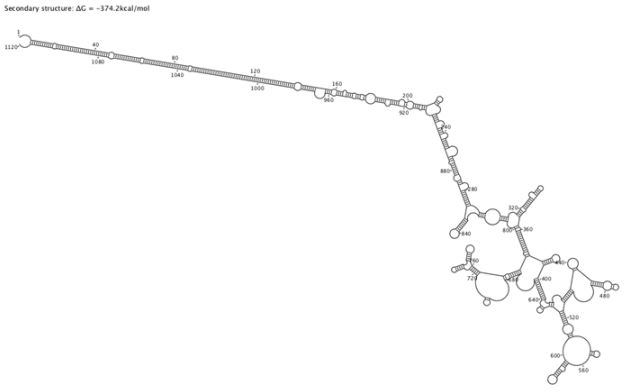


Figure 4.4. Repeat structure analysis of the extended 110kb *EPSPS* build. The top portion of the figure shows the direct and inverted repeats in an arc diagram format with the *EPSPS* gene annotated in yellow. Direct repeats are indicated by orange, red, and grey arcs while inverted repeats are indicated by green and blue. The bottom part of the figure shows the same repeat structures, but as a more conventional dot plot. Colors in the dot plot correlate to the colors of the annotated structures on the sequence seen between these two diagrams (red 2kb direct repeats).

5' Long Inverted Repeat (Putative MITE)

TAAGATTAGGAAAAATTGTCGGGAATAATCCGAACTATTTGCAAACTGCT
GTGAATAATCCCACGTATTGATTATTTATGAATAATCCCACCTTTCAAGTG
TATTTGCTCGTGGCACCCCCAAATGAAATTTGACCTATTTTAGTAGGTTA
TCTTCTTCAATGTCTTCTTCAATGCCTCTTTATAAACCCAGCTACTGATT
TGTATCCCACAAGCCATTGTTCTTCTTCAATTTATTCCACTTTGTTCTTC
AATCTTCACCTTTCTTCTTCCATTGTGTTCTTCCTTCTTCACTATTAACCC
TACGCAAGCCCTCTTCAAATGTATTACAATTTTGAATCAAATAATACAAT
TGATGCTCATAATTAACACCAAGACTAGTGACCACCAAATCATTAAGATC
AAACCATGAAATGCAATCAGGATCAAGTGAAAGGCTTCTATATTCCCCAC
CCACATAATTCAACCCTACCCCAGTCCTTTTGAATTTACCCCCATACCAAA
ACATCACTTGAAATTTTTCAAAATTATTAACCTAAAAAAACAACACAATT
GAACATAATTACCAATGCATTTCTATAACAACAAAGAAAACATTAAAGAA
TCAAAGATTAAAGTGAGGAATGGCAAAGAAATTACCATGGTTTGATTGA
GAACAAGAAGACCCAAAATTCGTCTGCACAGCCCCAAAATTTTCGCACA
GAGCAGCAATACCACCCCCAAAATTCGACACTGTTGATAAAAAATAAACC
CTAATTTTTTTGGGAAATTACAGTTGATGAATGTGAGTGTTGATTATGGC
GTGAAGCTTGATGATTATGAATGACAATTGTGCTTCAAGTTTTTGAAAT
TTTGAAGTTTTGAAGGAAGATGGTGTGAAGGAATGGTAGAACAGGAAAT
GAAGTTAAGGGTATGCCTTTTTGGGTTGAATGTTTATTTTATGGAATTAA
AGAATATGAAAGATCATACTCTAACCTGCAATAGTAGGTCAAATTTCATT
TGGGGGTGCCACGAGCAAATACACTTGAAAGGTGAGATTATTCATAAATA
ATCAATACTTGGGATTATTCACATAGGTTTGCGAATAGTTCGGATTATTC
CCAACAATTTTTCCTTAAGAT



3' Long Inverted Repeat (Putative MITE)

GACAATGAACTTTTTTTTTTAAAACTTTTTTTAAAAATATTTGGCTTT
TTAAGAGCTCCCTCAAATTATTTGCAAAAATAAGCTTCTACGCATAATGG
GCTGGGGGCTGGGTTTTGAGCTGGTAATGGGCCGCTGTTAGTACAAGCG
GACAATTCATTATGGGCTGATTTACCAATCGGCCGCTGTTTATATCAGCG
GGCCATTGGTCTTCATGGAATGGGCCGCATTCGGCCGCTGTTTGTATCAG
CGGGCCATTGGGCTTTCAATTAAAAACATTAATTTTTCAAAAAATGTGA
ATGTTTACCTCCTCCACTTCAGCTTCATCTTCTTCGCTGCTGCAATTCAT
GGTGATTGTGATTTGCCGCAATCAATGTTGCAGAGTAATAAGCACCGACC
ATGCAACGGAAACGAGAATGAGAAAATGTGATTTAATTCGTGATGCTTC
AAGTTCTAATTGTTGTCAACTTTGCATCAATTGCAGATTATTAACGACCG
AAATGACAATTGCAGCAATACGATCCTTATATTCAACAATGGAAGTTCAG
ATCAACTGTAGATGACTTGCCTGAAAATCAATTGAAGAACCAGGTTGAA
GATGAATTCGATCAACTTCGGAACAATTGGGGTTCCGTTATTTAACAGTA
AAACGTCCTGGGTCTGAAGCCCAATGGCCCGCTGATATAAACAGCGGCCG
AATGCGGCCCATTCCATGAAGACCAATGGCCCGCTGATATAAACAGCGGC
CGATTGGAAAATAGCCCCAAAATGAATTGCCCGCTTGTACTAACAGCGGC
CCATTACCAGCTCAAAACCCAGCCCCCAGCCCGTTACGCGTAGAAGCTTA
TTTTTGCAAATAATTTGAGGGAGCTCTTAAAAAGCCAAATATTTTTAAA
AAAAAGTTAAAAAAAAAAAAGTTCGACAAT

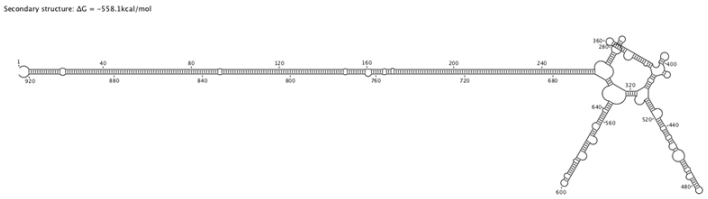
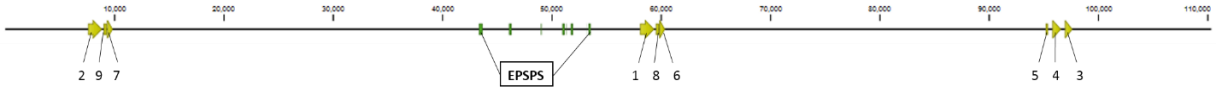


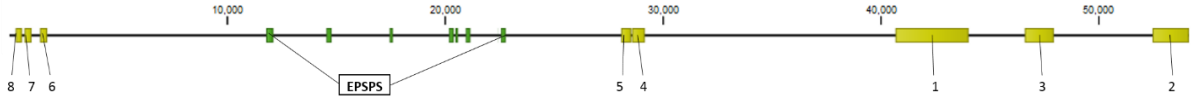
Figure 4.5. Sequence and secondary structure of the long inverted repeats (LIRs) flanking the *EPSPS* gene. Inverted repeats are indicated in red and tandem site duplications are underlined. LIR1 spans 41890-43009 in the extended *EPSPS* build and has 145bp-long nearly perfect inverted repeats just inside of the 6bp tandem site duplications. LIR2 spans 54982-55905 in the extended *EPSPS* build and has 260bp-long nearly perfect inverted repeats just inside of the 6bp tandem site duplications. The secondary structure of each LIR is shown to the right of the sequence.

A



				Top NCBI BLAST Hit				Putative Conserved Domain		
ORF #	Start	End	Length	Top BLAST Hit	Score	E-value	Identity	Putative Conserved Domain	E-value	Accession
1	58060	59370	1311	Transposase [Amaranthus palmeri]	846	0.0	100%	zf-BED (BED zinc finger)	2.81e-5	pfam02892
2	7552	8862	1311	Transposase [Amaranthus palmeri]	846	0.0	100%	zf-BED (BED zinc finger)	2.81e-5	pfam02892
3	95771	96577	807	Uncharacterized protein LOC104884087	161	2e-42	46%	PMD (Plant mobile domain)	5.99e-11	pfam10536
4	96913	97623	711	Uncharacterized protein LOC104884087	78	2e-13	38%	PMD (Plant mobile domain)	1.15e-4	pfam10536
5	94673	95056	384	Uncharacterized protein LOC104898784	94	2e-21	55%	zf-GRF (GRF zinc finger)	2.3e-4	pfam06839
6	59978	60352	375	Transposase [Amaranthus palmeri]	303	4e-96	99%	Dimer_Tnp_hAT (hAT family C-terminal dimerization region)	1.38e-22	pfam05699
7	9470	9844	375	Transposase [Amaranthus palmeri]	303	4e-96	99%	Dimer_Tnp_hAT (hAT family C-terminal dimerization region)	1.38e-22	pfam05699
8	59499	59801	303	Transposase [Amaranthus palmeri]	132	1e-33	95%	DUF4413 (Domain of unknown function)	3.2e-16	pfam14372
9	8991	9293	303	Transposase [Amaranthus palmeri]	132	1e-33	95%	DUF4413 (Domain of unknown function)	3.2e-16	pfam14372

B



				Top NCBI BLAST Hit				Putative Conserved Domain		
ORF #	Start	End	Length	Top BLAST Hit	Score	E-value	Identity	Putative Conserved Domain	E-value	Accession
1	40664	44017	3354	Gag-pol polyprotein [Silene latifolia]	1292	0.0	58%	RVT_1 (reverse transcriptase)	1.61e-37	pfam00078
2	52464	54131	1668	Uncharacterized protein LOC104902491	513	4e-164	46%	RVT_3 (reverse transcriptase-like)	1.41e-11	pfam13456
3	46597	47928	1332	UDP-D-glucuronate 4-epimerase 3 [Theobroma cacao]	791	0.0	88%	UDP_GE_SDE_e (UDP glucaronic acid epimerase)	0.0	cd05253
4	28571	29164	594	Protein FAR1-RELATED SEQUENCE 5-like [Beta vulgaris]	286	4e-89	65%	Protein FAR-RED ELONGATED HYPOCOTYL 3	3.5e-49	PLN03097
5	28076	28528	453	Protein FAR1-RELATED SEQUENCE 5-like [Beta vulgaris]	120	2e-28	46%	Protein FAR-RED ELONGATED HYPOCOTYL 3	1.1e-8	PLN03097
6	1392	1736	345	reverse transcriptase [Arabidopsis thaliana]	120	8e-29	54%	GAG pre-integrase domain	1.53e-14	pfam13976
7	685	1008	324	retrotransposon protein, Ty1-copia subclass [Oryza sativa]	124	6e-30	64%	No putative domain	-	-
8	275	577	303	polyprotein [Oryza australiensis]	94	2e-19	52%	RVT_2 (reverse transcriptase)	1.06e-4	pfam07727

Figure 4.6. Open reading frame (ORF) discovery and annotation for the extended *EPSPS* build (A) and the non-duplicated *EPSPS* fosmid consensus (B). ORFs are indicated by yellow arrows and numbered in order of length. Annotations from NCBI BLAST are shown for each ORF, including the putative conserved domain matching that sequence. The *EPSPS* gene is shown as green arrows in each build.

Figure 4.7. ClustalW alignment of the two EPSPS proteins in Palmer amaranth. Amino acid changes are in boxes.

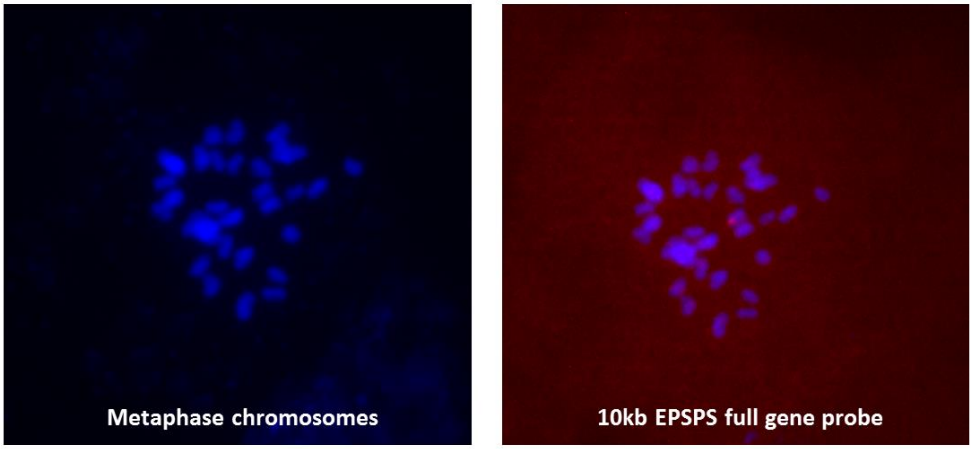


Figure 4.8. Fluorescent in situ hybridization of the full 10.2kb *EPSPS* gene against a glyphosate resistant Palmer amaranth genome. DAPI staining of the chromosomes are seen in the left figure and a merged image of the DAPI-stained chromosomes plus the Texas red labeled *EPSPS* probe is seen on the right.

REFERENCES

Altshuler D, Pollara VJ, Cowles CR, Etten WJV, Baldwin J, Linton L, and Lander ES. (2000) A SNP map of the human genome generated by reduced representation shotgun sequencing. Nature, 407: 513-516.

Devonshire AL and Field LM. (1991).Gene amplification and insecticide resistance. Annu Rev Entomol, 36: 1-21.

Eickbush TH, and Eickbush DG. (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics, 175: 477-485.

Feschotte C and Pritham EJ. (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet, 41: 331–368.

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang JM, Preston C, Leach JE, and Westra P. (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proc Natl Acad Sci USA, 107:1029–1034.

Gaines TA, Ward SM, Bukun B, Preston C, Leach JE, and Westra P. (2012) Interspecific hybridization transfers a previously unknown glyphosate resistance mechanism in *Amaranthus* species. Evol Appl, 5: 29-38.

Giacomini D, Westra P and Ward SM. (2014) Impact of genetic background in fitness cost studies: an example from Palmer amaranth. Weed Sci, 62:29-37.

Giacomini DA, Tao N, Ulmasov T, Latrielle P, Kerstetter R, Ward SM, Westra P, and Sammons D. (2015) Extending the EPSPS amplicon: steps towards defining the duplication mechanism. (218) Weed Science Society of America Meeting 2015, Lexington, Kentucky.

Gordon D, Abajian C, and Green P. (1998) Consed: A Graphical Tool for Sequence Finishing. Genome Res, 8: 195-202.

Gray YH. (2000) It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet, 16: 461-468.

Huang X, Wang J, Aluru S, Yang S, and Hillier L. (2003) PCAP: A Whole-Genome Assembly Program. Genome Res, 13: 2164-2170.

Kato A. (1999) Air drying method using nitrous oxide for chromosome counting in maize. Biotech Histochem, 3: 160-166.

Kato A Albert PS, Vega JM, and Birchler JA. (2006) Sensitive fluorescent in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotech Histochem, 81: 71-78.

Lander ES and Waterman MS. (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis, Genomics, 2: 231-239.

Leister D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet, 20: 116-122.

Nandula VK, Wright AA, Bond JA, Ray JD, Eubank TW, and Molin WT. (2014) EPSPS a mplification in glyphosate-resistant spiny amaranth (*Amaranthus spinosus*): a case of gene transfer via interspecific hybridization from glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*). Pest Manag Sci, 70:1902-1909.

Okonechnikov K, Golosova O, Fursov M, and the UGENE team. (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics, 28: 1166-1167.

Sammons, RD and Gaines, TA. (2014) Glyphosate resistance: state of knowledge. Pest Manag Sci, 70: 1367-1377.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, and Miller W. (2000) PipMaker – A web server for aligning two genomic DNA sequences. Genome Res, 10: 577-586.

Shaw CJ and Lupski JR. (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet, 13: R57-R64.

Shyr YYJ, Hepburn AG, and Widhohn JM. (1992) Glyphosate selected amplification of the 5-enolpyruvylshikimate-3-phosphate synthase gene in cultured carrot cells. Mol Gen Genet, 232: 377-382.

Szczesny P. (2008) Visualization of internal repeats in proteins (or DNA). Blog post, Freelancing science. GitHub repository, https://github.com/freesci/java-protein-domain-visualization.

Vila-Aiub MM, Goh SS, Gaines TA, Han H, Busi R, Yu Q, and Powles SB. (2014) No fitness cost of glyphosate resistance endowed by massive EPSPS gene amplification in Amaranthus palmeri. Planta 239:793-801.

Wattenberg M. (2002) Arc diagrams: visualizing structure in strings. In Proceedings of the IEEE Symposium on Information Visualization (InfoVis'02). IEEE Computer Society, Washington DC, 110.

Wiersma AT, Gaines TA, Preston C, Hamilton P, Giacomini D, Robin Buell C, Leach JE, and Westra P. (2015) Gene amplification of 5-enol-pyruvylshikimate-3-phosphate synthase in glyphosate-resistant Kochia scoparia. Planta, 241:463-474.

CHAPTER 5. CONCLUSION

In the Georgia Palmer amaranth populations studied here, the *EPSPS* gene was found to be duplicated in the genome more than 150 times. The duplicated *EPSPS* segment was sequenced and assembled out to 110kb, with more than 75% of the sequence flanking *EPSPS* consisting of long direct repeats, variable microsatellite regions, and transposable elements. For plants with 150 copies, this duplication added up to more than 1.5% of the total size of the Palmer amaranth genome (533Mbp/1C). Comparative analysis of this duplicated region with the non-duplicated *EPSPS* locus highlighted the presence of two MITEs at either end of the gene, marking the sites of divergence between the two assemblies. This likely indicated a transposon-mediated movement of the *EPSPS* gene into a unique genomic region. The dense repeat structure found at this new region along with preliminary fluorescent detection of the *EPSPS* gene supported a gene duplication mechanism based on unequal recombination. In this duplication model, the long 2kb direct repeats flanking the EPSPS amplicon would be prone to misalignment. When these misaligned regions recombine (during interphase in mitosis and/or prophase I in meiosis), the resulting daughter cells would end up with differing numbers of *EPSPS* gene copies.

This unequal recombination hypothesis was further supported by the inheritance studies which found a chimeric genotype for Palmer amaranth individuals with regards to *EPSPS* gene copy numbers. The previously documented complicated mode of inheritance was found to be at least partially caused by continually changing *EPSPS* copy numbers across the meristems of a single plant. With each floral meristem from a single female or male plant producing pollen or eggs with a range of gene copy numbers, the offspring

ended up with far fewer or more *EPSPS* gene copies than would be predicted by Mendelian segregation. Unequal recombination may also occur during meiosis, adding further variation in copy number to this system, but more work is needed to confirm this. The inheritance studies also found a maintenance of high *EPSPS* copies even in the absence of a glyphosate application selection pressure.

Intriguingly, this chromosomal instability did not appear to have a negative effect on the growth and survival of the Palmer amaranth plants in this study. When single copy *EPSPS* plants were grown up alongside multiple *EPSPS* copy plants and compared across every step of the growth cycle, no difference in fitness was found, either vegetatively or reproductively. This finding, combined with the inheritance results, indicates that this form of glyphosate resistance is unlikely to recede from the natural populations even if farmers stop using glyphosate in their management systems for a number of years. Longer term studies are needed to determine whether this pattern will persist for more than three generations, but the data presented here suggests a long-term loss of glyphosate from weed management plans for growers in areas with established glyphosate resistant Palmer amaranth.

On a brighter note, this work is exciting for the world of basic plant science. A case of single gene duplication on this scale is unusual in the plant kingdom and the rapid gain and loss of gene copies is even more so. Although the results reported here suggest an unequal recombination mechanism of *EPSPS* gene duplication after an initial transposition to a repeat rich chromosomal region, more studies are needed to confirm this result; it may be that the mechanism behind this duplication is completely novel. Elucidating the full genetic architecture and regulatory pathways behind this adaptive gene duplication would

be a significant step in learning more about the ways in which plants sense unusual events and restructure their genome in response to them.